

POLITECNICO DI MILANO
Corso di Laurea Magistrale in Mathematical Engineering
Dipartimento di Matematica

Scuola di Ingegneria Industriale e dell'Informazione



**Controllo Statistico di oggetti a geometria
complessa con applicazioni all'Additive
Manufacturing**

Relatore: Prof. Piercesare Secchi
Correlatore: Dott.ssa Alessandra Menafoglio

Tesi di Laurea di:
Riccardo Scimone, matricola 892143

Anno Accademico 2018/2019

A Gloria, Mamma, e Papà

Sommario

Il lavoro si colloca nell'ambito dello Statistical Process Control esteso a processi di Additive Manufacturing, in cui possono essere spesso prodotti oggetti con geometrie non banali e piuttosto articolate. Lo scopo della tesi è l'applicazione e lo sviluppo di metodi statistici adatti al controllo della produzione di oggetti di questo tipo, con una costante attenzione tanto alle basi teoriche su cui si fondano i metodi quanto all'applicazione pratica degli stessi. Si avrà occasione di approfondire ed apprezzare l'efficacia di strumenti provenienti da branche di matematica piuttosto avanzata (analisi in spazi di Hilbert, Geometria Differenziale), che si mostrano sorprendentemente potenti nell'analisi di oggetti anche molto complessi.

Ringraziamenti

Ringrazio di cuore il Prof. Piercesare Secchi, relatore di questo lavoro di tesi, per il grandissimo apporto di esperienza e conoscenze, ma soprattutto per aver trasmesso, insieme ad Alessandra, un'autentica passione per la Statistica, che mi ha consentito di affrontare questo lavoro con gioia ed entusiasmo. Ringrazio Alessandra per essere stata una correlatrice costantemente disponibile, estremamente competente e sempre col sorriso sulle labbra.

Un sentito ringraziamento va a tutti quelli che, rimasti in Sicilia, vedo ormai di rado: gli amici del liceo, i tantissimi compagni di squadra, e i componenti delle società PGS Domenico Savio e AS Pallavolo Messina. In particolare, rivolgo un pensiero speciale a Tonino, sempre disponibile per serate divertenti durante le vacanze, e Flavio, che è stato per me un vero maestro di pallavolo. Nei cinque anni trascorsi in Lombardia, ho avuto modo di stringere amicizia con tantissimi colleghi, compagni di squadra, e compagni di residenza. Per questo ringrazio tutti i ragazzi del Campus Martinitt, che per tre anni sono stati una seconda famiglia, ed i ragazzi della Casa dello Studente: Sara, Alessandra, Mimmo, Chiara, Mattia, Gianfranco, Marco, Leonardo, Thomas, e tantissimi altri.

In particolare, ringrazio con tutto il cuore Dario, in cui ho trovato un allegro compagno di camera, di vacanze, di pasti e di bevute, ma soprattutto un amico fraterno e un esempio da seguire.

Rivolgo un pensiero anche a tutte le squadre per cui ho giocato in Lombardia, ovvero le società di Desio, Segrate e del Gonzaga. Ringrazio soprattutto i miei compagni Buzzo e Mario, per la grande presenza di spirito durante ogni allenamento, mentre Simone dovrebbe invece ringraziare noi tre per averlo allenato con tanta abnegazione.

Un ringraziamento speciale va alla Zia Cettina e allo Zio Saro, che mi hanno accolto ed aiutato come un figlio da ormai cinque anni, ed alla cugina Sabrina. Al termine di questo percorso lungo, faticoso e ricco di soddisfazioni, il più grande ringraziamento è rivolto alla mia famiglia, nonni compresi, costante fonte di fiducia e di supporto morale e materiale, porto sicuro nei momenti di gioia e in quelli di sconforto, nonostante i molti chilometri di distanza. Mi avete guidato e fatto crescere circondato di amore, e niente potrà mai ripagarvi.

Indice

Sommario	3
Ringraziamenti	5
1 Introduzione	9
2 Dati, forme e rappresentazioni: introduzione al dataset oggetto di studio	13
2.1 Superfici, Point Clouds e mesh	13
2.2 Il problema della rappresentazione	14
2.3 Illustrazione del dataset oggetto di studio e software per il processamento delle mesh	14
3 Distanza di Hausdorff e rappresentazione delle mesh	19
3.1 Confronto tra mesh e distanza di Hausdorff	19
3.1.1 Definizioni e proprietà di base	19
3.1.2 Significato geometrico e asimmetria delle mappe di distanza	20
3.1.3 Il problema dell'allineamento	21
3.2 Formulazione in termini di densità di probabilità	22
3.2.1 Estensione ad un dataset di oggetti e proprietà della rappresentazione	24
4 Spazi di densità di probabilità e SFPCA	27
4.1 Dati composizionali	27
4.1.1 Geometria di Aitchison e operazioni nel simpleso . .	29
4.2 Spazi di Hilbert	34
4.2.1 Prime definizioni e proprietà principali	34
4.2.2 Osservazioni e PCA in spazi di funzioni	37
4.3 Lo spazio B^2 e le sue caratteristiche	38
4.3.1 Somma e prodotto	39
4.3.2 Spazi di Banach di densità	40
4.3.3 Lo spazio B^2	44
4.4 PCA nello spazio B^2	46

4.4.1	Possibili carte di controllo derivanti da SFPCA e conclusioni	48
5	Caso studio - Prima parte	51
5.1	Fase preliminare	51
5.1.1	Codifica, pulizia e allineamento delle mesh	51
5.1.2	Calcolo delle mappe, stima delle densità e riduzione a un supporto unico	52
5.2	Analisi del dataset \mathbf{f}_P	54
5.2.1	Ispezione del dataset e stima delle densità	55
5.2.2	Calcolo delle componenti principali e tipologie di controllo	59
5.2.3	Analisi della varianza spiegata, proiezioni, interpretazione delle autofunzioni	60
5.2.4	Carta T^2 in Fase 1 e 2, possibili adattamenti, conclusioni	63
5.3	Analisi del dataset \mathbf{f}_S	65
6	Operatore di Laplace-Beltrami e analisi di superfici	69
6.1	Introduzione e cenni di geometria riemanniana	69
6.1.1	Prerequisiti	71
6.1.2	Fondamenti di geometria riemanniana, connessione di Levi-Civita	72
6.1.3	Calcolo integrale e operatori differenziali su varietà riemanniane	75
6.2	Proprietà dell'operatore	78
6.3	Soluzioni per il calcolo dello spettro	82
7	Caso studio - Seconda Parte	85
7.1	Un nuovo strumento nel Controllo Statistico di Processo . . .	85
7.2	Analisi del dataset oggetto di studio	86
8	Conclusioni e sviluppi futuri	97

Capitolo 1

Introduzione

I processi di Additive Manufacturing (AM) diventano sempre più importanti, sia per via di un continuo miglioramento delle tecnologie che ne sono alla base, sia per il gran numero di geometrie realizzabili attraverso tali processi. La possibilità di produrre oggetti con geometrie complesse (ad esempio, con completa assenza di simmetrie, strutture reticolari interne non regolari) costituisce una grande sfida per la disciplina del Controllo Statistico di Processo (o SPC, da Statistical Process Control), il cui scopo primario è il monitoraggio dei processi produttivi, volto all'individuazione di eventuali oggetti non conformi agli obiettivi di produzione e a determinare, di conseguenza, se il processo è “in controllo” (produce sistematicamente pezzi di qualità accettabile) o “fuori controllo” (produce in maniera sistematica oggetti non accettabili rispetto agli obiettivi di produzione). La necessità di sviluppare metodi di controllo adeguati è inoltre sostenuta dal fatto che i settori in cui i processi AM risultano più ricchi di potenzialità sono, tra gli altri, l'aerospaziale ed il biomedicale, aree industriali in cui la prevenzione e l'individuazione dei difetti di produzione risultano fondamentali (Grasso e Colosimo 2017), tanto per questioni economiche quanto, soprattutto, per la gravità delle possibili conseguenze derivanti dalla mancata individuazione di uno scostamento dagli standard produttivi. Nel settore aerospaziale, in particolare, esistono casi in cui la produzione AM presenta peculiarità che rendono un adeguato controllo di processo assolutamente imprescindibile: si pensi ad una situazione di produzione in ambiente extraterrestre, in cui è necessario rispettare strettissimi vincoli dovuti alla limitata quantità di energia e di materie prime disponibili, spesso appena sufficienti alla produzione di un unico esemplare che, di conseguenza, *deve* essere conforme agli standard. La necessità di progressi scientifico-tecnologici in questa ed altre aree collegate al settore spaziale ha spinto il Politecnico di Milano e l'Agenzia Spaziale Italiana ad un accordo di collaborazione scientifica pluriennale. Tale accordo, imperniato su diverse linee di ricerca scientifica e tecnologica, presenta tra gli obiettivi principali un concreto miglioramento nelle prestazioni e nel

monitoraggio dei processi AM¹.

Lo scopo della tesi è dunque l'applicazione e lo sviluppo di metodi di controllo statistico che siano applicabili ad oggetti con geometrie non standard, con particolare attenzione al caso in cui si desideri ottenere una misura dello scostamento degli oggetti prodotti rispetto ad un prototipo nominale, o al caso in cui si vogliano misurare le differenze di un nuovo oggetto rispetto ad un campione di produzione che si suppone essere a priori in controllo. Il lavoro è essenzialmente imperniato su tre fulcri teorici: la distanza di Hausdorff tra insiemi, la Simplicial Functional Principal Component Analysis (SFPCA, Hron et al. 2014), e l'operatore di Laplace-Beltrami per varietà Riemanniane.

I primi due elementi risulteranno saldamente collegati tra loro, in quanto mostreremo come le mappe di distanza ottenute dalla misura della distanza di Hausdorff tra due insiemi possano essere facilmente ricondotte a densità di probabilità, oggetti la cui variabilità è studiata nella SFPCA. La distanza di Hausdorff è uno strumento ormai consolidato nello studio delle differenze tra *mesh*, ovvero ricostruzioni poliedriche, spesso a facce triangolari, di oggetti tridimensionali. Risulta essere un metodo naturale e intuitivo, e per questo motivo un gran numero di lavori si sono dedicati a costruirne varianti o a rendere più semplice e più efficiente il calcolo della distanza di Hausdorff tra oggetti (Cignoni et al. 1998; Aspert et al. 2002; Guthe et al. 2005; Oniga e Chirila 2013; solo per citarne alcuni), al punto da rendere possibile in tempi accettabili anche il calcolo delle distanze su *mesh* molto grandi e complesse. L'applicazione della SFPCA all'output ottenuto dalla misura delle distanze permetterà di raggiungere un notevole livello di interpretabilità dei risultati, sia per via del preciso significato geometrico delle distanze, sia per via della grande quantità di lavoro teorico svolto nell'ambito dell'analisi di dati distribuzionali (i.e., densità di probabilità) (Egozcue et al. 2006; Pawlowsky-Glahn et al. 2014; Hron et al. 2014). L'uso della SFPCA ha già trovato recente applicazione alla produzione industriale nel monitoraggio della porosità superficiale di schiume metalliche attraverso l'analisi di immagini (Menafoglio et al. 2018), un tipo di dato che è sempre più presente nei processi di controllo industriale (Yan et al. 2015; Huang et al. 2015). Tuttavia, l'applicazione di tale metodo a densità di probabilità provenienti dal calcolo delle distanze tra *mesh* sembra essere un problema nuovo, su cui la tesi si concentrerà per tutta la prima parte.

L'operatore di Laplace-Beltrami su varietà risulterà invece il fulcro della seconda parte della tesi, in cui verranno mostrate le proprietà essenziali di tale operatore, il cui spettro risulta essere un potente strumento nell'analisi di superfici geometriche anche piuttosto complesse e, in particolare, uno strumento *intrinseco*, ossia dipendente solo dalle proprietà geometriche della

¹Si consulti https://www.polimi.it/fileadmin/user_upload/comunicati_stampa/1516894344_Comst_ASI_POLIMI.pdf per maggiori informazioni.

superficie da cui deriva. L'uso dell'operatore di Laplace-Beltrami nell'analisi di immagini e nella computer graphics è piuttosto consolidato, e trova radici nel pionieristico lavoro (Reuter et al. 2006), in cui lo spettro dell'operatore viene appunto proposto come “Shape-DNA”, ossia come impronta di forma geometrica caratteristica, sulla base di precedenti ricerche nell'ambito della teoria della geometria differenziale (Kac 1966; Gordon et al. 1992; Levy 2006). A questo lavoro se ne sono aggiunti molti altri, che hanno proposto varianti o generalizzazioni dell'operatore o, più spesso, metodi alternativi di calcolo dello spettro via via più rapidi ed efficienti (Belkin e Niyogi 2008; Li et al. 2015; Patané 2014; Patané 2017). Va però sottolineato che l'applicazione di simili tecniche spettrali allo Statistical Process Control industriale è estremamente recente, e costituisce un punto di congiunzione tra i mondi del controllo di produzione a quello del processamento di immagini e della computer vision (Zhao e Del Castillo 2019). Questi ultimi autori propongono le prime carte di controllo basate sullo spettro del Laplaciano. Sarà di interesse per il lavoro di tesi studiare l'efficacia di tale strumento all'analisi di *mesh* provenienti dalla tomografia di oggetti complessi prodotti via AM.

Ogni sezione della tesi sarà accompagnata da esempi di applicazione concreta ad oggetti reali prodotti dal gruppo di ricerca diretto dalla Prof. Bianca Colosimo al Dipartimento di Meccanica del Politecnico di Milano, la cui collaborazione ha reso possibile ottenere una misura reale dell'efficacia dei vari metodi che verranno illustrati. La tesi è articolata secondo la seguente struttura:

- Nel secondo capitolo si discute del concetto di mesh in maniera più dettagliata, si introduce il lettore al dataset di oggetti prodotti menzionato precedentemente, e viene affrontato il problema di rappresentare gli oggetti al centro dell'analisi in maniera semplice ma al tempo stesso significativa ed interpretabile.
- Nel terzo capitolo si illustrano i concetti fondamentali riguardanti la distanza di Hausdorff e la sua interpretazione geometrica, e si mostrano i passi necessari per trasformare un campione di mesh in una rappresentazione basata su densità di probabilità che hanno per supporto opportuni intervalli derivanti dal calcolo delle distanze di Hausdorff.
- Nel quarto capitolo, a carattere prevalentemente teorico, si percorre in maniera sintetica la formalizzazione matematica alla base della SFPCA, con particolare attenzione alla costruzione dello spazio di Hilbert B^2 di densità di probabilità, che risulterà fondamentale nell'applicazione pratica che sarà oggetto della quinta sezione.
- Nel quinto capitolo viene dato ampio spazio all'applicazione della SFPCA alle distribuzioni di distanza calcolate sul dataset a nostra

disposizione, e si mostrano pregi e difetti del metodo, con un'enfasi particolare sull'interpretabilità dei risultati prodotti dall'analisi. Tale capitolo, oltre che un esempio applicativo concreto, fornisce anche una dettagliata descrizione dei problemi pratici che si riscontrano nel processamento di un dataset di mesh, e delle possibili soluzioni a tali problemi.

- Nel sesto capitolo il lettore viene introdotto alle proprietà fondamentali dell'operatore di Laplace-Beltrami. Ci si concentrerà con maggiore attenzione sulle motivazioni che ne giustificano l'adozione come "DNA" della forma di una superficie, mostrando come dall'analisi delle proprietà spettrali di tale operatore sia possibile costruire una rappresentazione di un campione di mesh alternativa alla precedente, con annessi pro e contro.
- Nel settimo capitolo sono mostrati i risultati relativi al calcolo dello spettro per le *mesh* relative al dataset a disposizione, e vengono condotti gli opportuni confronti con il metodo proposto nella prima parte della tesi.
- Nell'ottavo e ultimo capitolo vengono brevemente delinate alcune conclusioni sull'ambito di applicazione dei metodi illustrati, come anche alcuni confronti sui punti di forza e debolezza dei metodi stessi.

Capitolo 2

Dati, forme e rappresentazioni: introduzione al dataset oggetto di studio

2.1 Superfici, Point Clouds e mesh

Come fatto notare da Zhao e Del Castillo 2019, l'industria manifatturiera moderna ha visto crescere in maniera tumultuosa la qualità e la quantità di tecnologie in grado di scansionare oggetti tridimensionali complessi, in maniera sempre più rapida, precisa, affidabile e, ovviamente, economica. Tali avanzamenti tecnologici hanno avuto, come immediata conseguenza, l'aumento della complessità dei dataset coinvolti in analisi di Controllo Qualità e Controllo Statistico di Processo. In particolare, il risultato della scansione/tomografia di una parte meccanica prodotta risulta tipicamente essere una rappresentazione, più o meno sintetica e semplificata, di una superficie bidimensionale immersa nell'usuale spazio euclideo (si potrebbe anche parlare di varietà riemanniana, ma non avremo bisogno di formalizzare concetti di geometria differenziale fino al capitolo 5). L'output più semplice di una tomografia è, di norma, una *Point Cloud*, ovvero un insieme di diverse migliaia o decine di migliaia di punti nello spazio, ognuno caratterizzato da un vettore \mathbf{x} di coordinate, giacenti su una superficie ignota, a meno di una certa tolleranza tipica del tomografo o di qualunque altro strumento utilizzato per la scannerizzazione. È piuttosto comune che un tomografo sia anche in grado di produrre triangolazioni della *Point Cloud* registrata, ovvero poliedri a facce triangolari (ma a volte anche quadrate) che, a seconda dell'accuratezza

e della quantità dei punti registrati, forniscono un'approssimazione più o meno precisa dell'oggetto reale o della parte meccanica di partenza. Per tutta l'estensione della tesi, per *mesh* si intenderà dunque un poliedro a facce triangolari, rappresentazione di un oggetto reale, e non ci si occuperà, in generale, di quanto tale rappresentazione sia precisa: le *mesh* saranno il punto di partenza fondamentale di tutte le analisi che verranno condotte.

2.2 Il problema della rappresentazione

Volendo mantenere un approccio il più possibile generale e vicino alla realtà, è evidente che occorre pensare una *mesh* come una superficie bidimensionale immersa nello spazio euclideo. Pertanto, un campione di N esemplari prodotti si traduce in un campione di N superfici, tra loro diverse per via della variabilità di processo. Al fine di renderne possibile l'analisi, è tuttavia necessario costruire una rappresentazione alternativa, e possibilmente semplificata, che conservi tutte le informazioni che si ritengono significative ai fini del monitoraggio di processo, ma che trasformi le superfici in enti matematici che siano più semplicemente trattabili. In particolare, sarebbe opportuno che il dataset ottenuto a valle dell'operazione di rappresentazione possa essere immerso in un opportuno spazio vettoriale che renda possibile l'applicazione delle comuni tecniche di analisi. È chiaro come entrambi i requisiti siano difficilmente raggiungibili se si decide di lavorare direttamente su un dataset i cui atomi sono superfici, e ciò lascia aperto il problema di scegliere una rappresentazione adeguata dei dati, problema per il quale non esiste una soluzione corretta o univoca, ma diverse soluzioni più o meno efficaci e ragionevoli. In questo lavoro, come già anticipato nell'introduzione, verranno presentate in dettaglio due possibili soluzioni, che sono state ritenute opportune per via di motivi che diverranno chiari proseguendo nella lettura, e ciascuna soluzione presenta diversi vantaggi e svantaggi, che verranno analizzati a tempo debito. Ciò che si desidera sottolineare in questa sezione è che la scelta di una rappresentazione costituisce un passo di capitale importanza all'interno dell'attività di analisi, un passo che influenza inevitabilmente la qualità, il significato e l'interpretabilità di qualunque risultato raggiunto. Per questi motivi, si cercherà sempre di illustrare in maniera il più dettagliata e completa possibile le proprietà delle diverse rappresentazioni scelte. Forniamo adesso una panoramica delle caratteristiche del dataset a nostra disposizione.

2.3 Illustrazione del dataset oggetto di studio e software per il processamento delle *mesh*

Il dataset su cui tutte le analisi sono state svolte consiste in un campione di 16 *mesh*, risultanti dalla tomografia di oggetti di materiale plastico prodotti

via Additive Manufacturing dal gruppo di ricerca diretto dalla Prof. Bianca Colosimo presso il Dipartimento di Meccanica del Politecnico di Milano. Non si tratta di parti meccaniche reali o di oggetti di effettiva utilità pratica, in quanto si è preferito concentrarsi, nell'ambito di questo lavoro, su oggetti che potessero fornire un livello di complessità geometrica tale da giustificare l'utilizzo di tecniche di controllo statistico originali ed avanzate, e che, d'altro canto, fossero producibili con tempi e costi ragionevoli. Si tratta dunque di *uova cave*¹, caratterizzate da una struttura trabecolare irregolare, di cui mostriamo un esempio in Figura 1. Si noti che, oltre alla struttura irregolare dei trabecoli, quasi irriproducibile da qualunque processo produttivo che non sia AM, gli oggetti prodotti sono anche caratterizzati da una curvatura superficiale e da una densità trabecolare piuttosto variabile. Si tratta dunque di un esempio decisamente realistico della complessità geometrica realizzabile nei processi di Additive Manufacturing industriale.

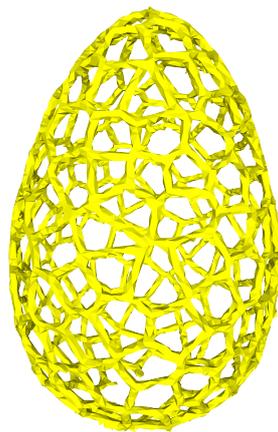


Fig.1, primo esempio degli oggetti con cui lavoreremo

Tutti e 16 gli elementi del dataset sono stati prodotti sulla base di un prototipo nominale comune che, nella prima parte del lavoro, sarà il riferimento fondamentale rispetto a cui verranno calcolate le distanze tra mesh. Per testare l'efficienza dei metodi che proporremo, sono stati introdotti dei difetti, a livelli di severità più o meno alti, su alcuni elementi del dataset. Si fornisce di seguito una classificazione degli elementi del dataset, accompagnata

¹Progettate dalla designer Giorgia Galimberti e realizzate attraverso stampa Fdm presso il laboratorio AddMe.Lab del Dipartimento di Meccanica del Politecnico di Milano

dalla nomenclatura adottata in tutto il seguito del lavoro e da alcune immagini illustrative.

- I primi 13 elementi del dataset sono considerati "in controllo", ovvero prodotti sulla base nominale in maniera accettabile rispetto all'obiettivo di produzione. Ci si riferirà a tali elementi con la dicitura IC_j , $j = \{1, \dots, 13\}$, dove IC sta per "In Control". Un esempio di questi elementi è riportato in Figura 1.
- Il quattordicesimo elemento del dataset è fuori controllo per lievi irregolarità distribuite su tutta la superficie. Ci si riferirà a questo elemento con OOC-LS, ovvero "Out Of Control, Low Severity".

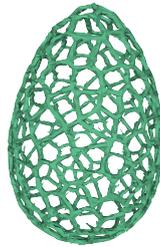


Fig.2, immagine di OOC-LS, si notino le lievi irregolarità

- Il quindicesimo componente è fuori controllo per irregolarità simili al precedente, ma più pronunciate ed evidenti. La sigla adottata per questo elemento sarà OOC-HS, ovvero "Out Of Control, High Severity".

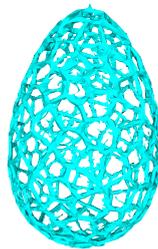


Fig.3, immagine di OOC-HS, notare le irregolarità molto evidenti

- Il sedicesimo e ultimo elemento è del tutto simile agli elementi in controllo, eccetto per la mancanza di un trabecolo. Ci riferiamo a

questo oggetto con OOC-MS, ossia "Out Of Control because of a Missing Strut"

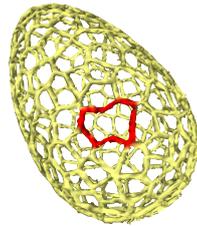


Fig.4, immagine di OOC-MS, ruotato in modo da apprezzare l'effettiva mancanza del trabecolo

Tutto il lavoro di processamento, pulitura, e confronto delle mesh è stato svolto grazie ai pacchetti software *Meshlab* (Cignoni et al., 2008; Cignoni et al. 1998) e *rvcg* (Schlager, 2017, Capitolo 9), entrambi interfacce della grande e consolidata libreria C++ *VCG*, la cui documentazione completa è consultabile al link vcg.isti.cnr.it/vcglib/index.html.

Capitolo 3

Distanza di Hausdorff e rappresentazione delle mesh

3.1 Confronto tra mesh e distanza di Hausdorff

3.1.1 Definizioni e proprietà di base

Mantenendo un contesto totalmente generale, si supponga di avere a disposizione due mesh, magari con delle piccole differenze come negli esempi visti in precedenza, e di volere costruire un indicatore che fornisca una quantificazione di quanto le mesh sono tra loro “distanti”. Introduciamo dunque uno strumento matematico utile a misurare distanze tra sottoinsiemi di uno spazio metrico, la distanza di Hausdorff.

Definizione 3.1.1. *Sia U uno spazio metrico con metrica d , e \mathcal{C} la famiglia di sottoinsiemi di U*

$$\mathcal{C} := \{C \subset U \text{ t.c. } C \text{ limitato, chiuso, non vuoto}\}$$

Chiamiamo allora distanza di Hausdorff la metrica d_H su \mathcal{C} definita da

$$d_H(X, Y) := \max \left\{ \sup_{x \in X} \inf_{y \in Y} d(x, y), \sup_{y \in Y} \inf_{x \in X} d(x, y) \right\}, \forall X, Y \in \mathcal{C}$$

Per comprendere più a fondo la definizione, possiamo definire le mappe $d_X : Y \rightarrow \mathbb{R}^+$, $d_Y : X \rightarrow \mathbb{R}^+$ tali che

$$d_X(y) := \inf_{x \in X} d(x, y), \quad d_Y(x) := \inf_{y \in Y} d(x, y)$$

La mappa d_X , ad esempio, associa ad ogni $y \in Y$ la sua distanza da X , e d_Y ha un ruolo esattamente analogo. È evidente che le mappe d_X e d_Y , se entrambe note, costituiscono una imponente e puntuale fonte di informazioni sulle differenze tra due insiemi, molto più ricca del semplice numero $d_H(X, Y)$

che, chiaramente, può andare incontro a variazioni estremamente significative quando anche un solo punto viene aggiunto a uno dei due insiemi. Chiamando poi $d_{XY} := \sup_{x \in X} d_Y(x)$, $d_{YX} := \sup_{y \in Y} d_X(y)$, si trova che d_{XY} è quel valore di distanza corrispondente al punto di X situato alla massima distanza da Y , e viceversa per d_{YX} (stiamo implicitamente supponendo che gli estremi superiori e inferiori siano in realtà massimi e minimi, senz'altro vero se le mappe sono continue e se X e Y sono compatti). Non è difficile convincersi del fatto che in generale $d_{XY} \neq d_{YX}$, fatto che approfondiremo nella prossima sezione, e che motiva la definizione $d_H(X, Y) := \max \{d_{XY}, d_{YX}\}$. In effetti, perchè una distanza sia tale è necessario che essa soddisfi la proprietà di simmetria. Si riportano, per completezza, alcuni risultati fondamentali relativi alla distanza di Hausdorff, utili per comprenderne più a fondo alcune proprietà matematiche. Per la dimostrazione si rimanda, ad esempio, a (Henrikson 1999).

Teorema 3.1.1. *Valgono i seguenti risultati.*

1. d_H è una distanza.
2. Se (U, d) è totalmente limitato, allora (\mathcal{C}, d_H) è totalmente limitato.
3. Se (U, d) è completo, allora (\mathcal{C}, d_H) è completo.

Molto lavoro è stato svolto nell'indagare la relazione tra la topologia indotta su \mathcal{C} dalla metrica d_H e la continuità delle misure sulla σ -algebra generata da \mathcal{C} , nonché nello studio della relazione tra d_H e gli insiemi frattali (ibid.). Non ci si addentrerà in questi campi, che esulano totalmente dagli argomenti oggetto della tesi.

3.1.2 Significato geometrico e asimmetria delle mappe di distanza

Proviamo adesso a fornire un'interpretazione geometrica delle mappe e degli indicatori introdotti nella precedente sotto-sezione. Concentriamoci sull'asimmetria $d_{XY} \neq d_{YX}$. Ci si restringa a $U = \mathbb{R}^3$, e supponiamo che $X = S^2$ sia la superficie sferica di \mathbb{R}^3 con raggio unitario e che $Y = S^2_+$ sia la semisuperficie sferica di raggio unitario giacente al di sopra del piano $z = 0$. Si trova immediatamente che $d_{YX} = 0$, in quanto ogni punto della semisfera appartiene anche alla sfera, pertanto $d_X(y) = 0 \forall y \in Y$. D'altro canto, $d_{XY} = \sqrt{2}$, poichè il punto della superficie sferica X a massima distanza da Y è il polo sud della sfera, a cui corrisponde appunto un valore $d_{XY} = \sqrt{2}$. E' dunque possibile fornire una prima concreta interpretazione geometrica dell'asimmetria delle mappe d_X e d_Y . Tale asimmetria è, in primo luogo, causata dal fatto che le due mappe hanno domini diversi: nel caso del calcolo di d_{YX} , ad esempio, per ogni punto di Y , ossia il dominio di d_X , si cerca il punto di X più vicino, e tale ricerca, pur esaurendo tutti i punti di Y , non

necessariamente esaurisce i punti di X , esattamente come nel caso dell'esempio appena illustrato. Nel caso di confronto tra mesh corrispondenti a oggetti reali, va dunque tenuto in conto che le mappe d_X e d_Y derivanti dal confronto trasportano informazioni che sono spesso complementari, e vanno dunque analizzate entrambe. Per fissare le idee, si pensi al dataset introdotto nella sezione precedente, e siano allora X il prototipo nominale e Y l'elemento OOC-MS, ovvero l'elemento fuori controllo per via della mancanza di un trabecolo. Si supponga di allineare X e Y , idealmente in modo perfetto, ovvero in modo tale che l'unica differenza rilevabile sia l'assenza del trabecolo. Si comprende immediatamente che, a questo punto, la mappa d_Y , avente per dominio il prototipo nominale, "vedrà" dei punti di X che non hanno corrispondenti in Y , e trasporterà pertanto le informazioni relative al difetto che speriamo di rilevare, mentre la mappa d_X ignorerà completamente tale difetto. Se invece l'elemento fosse difettoso per via di *abbondanza* di materiale, e.g., a causa di strutture trabecolari che dovrebbero essere assenti, tutte le considerazioni precedenti vengono ribaltate. Ribadiamo dunque che, in un confronto tra mesh, lo studio delle mappe di distanza può non solo rivelare *se* sono presenti delle anomalie, ma anche fornire concreti indizi sul *tipo* di anomalie che vengono rilevate, i.e., eccesso o difetto rispetto al nominale.

3.1.3 Il problema dell'allineamento

Va tenuto in conto che, qualunque sia lo strumento di calcolo utilizzato per ottenere la mappe di distanza sulle mesh, tali mappe hanno senso solo se calcolate successivamente ad un opportuno allineamento delle mesh stesse, che ne massimizzi la sovrapposizione. Nel caso del nostro dataset, questo fatto è assolutamente evidente: ogni confronto grafico può essere condotto solo se le strutture trabecolari degli oggetti che vengono confrontati sono massimamente sovrapposte. Si tratta di un particolare non indifferente, in quanto il processo di allineamento risulta spesso *costoso*, e non necessariamente dal punto di vista computazionale. Il problema di allineare *Point Clouds* simili è infatti ben noto in letteratura, e viene spesso risolto dall'algoritmo ICP, Iterative Closest Point (Besl e McKay 1992), che costituisce ormai uno standard affermato e relativamente efficiente (Zhao e Del Castillo 2019). Il *costo* effettivo è dovuto al fatto che gli algoritmi di allineamento necessitano di un ragionevole livello di allineamento spaziale iniziale per giungere a convergenza, cosa che richiede nella quasi totalità dei casi la presenza di un operatore umano che imprima alle mesh delle opportune trasformazioni rigide: questo è esattamente il caso dell'affermato software *Meshlab*, usato per manipolare il dataset a disposizione. In un'ottica di controllo statistico in produzione, è evidente che la necessità continua di un intervento umano nel processo che va dalla tomografia degli oggetti all'effettiva analisi statistica delle mesh potrebbe costituire un ostacolo non indifferente, specie in un contesto di controllo *on-line*, ossia in tempo reale. Questo ostacolo viene rimosso dall'uso

di proprietà *intrinseche* delle superfici, come fatto notare in (Zhao e Del Castillo 2019), aspetto su cui ci concentreremo a partire dalla quinta sezione. Fino alla fine della quarta sezione, in ogni caso, si intenderà sempre che gli oggetti/insiemi/mesh di cui vengono calcolate le mappe di distanza siano stati debitamente sovrapposti e allineati attraverso opportune rototraslazioni nello spazio euclideo. Passiamo adesso ad una breve descrizione dell’algoritmo ICP, implementato da Meshlab, che è stato usato per allineare gli elementi del dataset oggetto di studio. Per una descrizione completa si rimanda a (Besl e McKay 1992) in cui sono anche presentati risultati di convergenza e varianti accelerate dell’algoritmo. Sia P la figura di riferimento, ovvero una *Point Cloud*, una mesh, o una superficie parametrica. P costituisce il riferimento spaziale rispetto a cui effettuare l’allineamento, per cui il suo orientamento non viene alterato dall’esecuzione dell’algoritmo. Sia X l’ente geometrico che subisce l’allineamento. X deve essere una *Point Cloud* o comunque approssimato da un numero N finito di punti. Nel caso in cui sia P che X siano due mesh, tipicamente X viene identificata con la *Point Cloud* costituita dai suoi vertici. L’algoritmo è articolato come segue:

1. Si pone $k = 0, X_0 = X$.
2. Alla k -esima iterazione, per ogni punto $x_j \in X_k, j = 1, \dots, N$ si trova il punto $p_j \in P$ più vicino a x_j . Tali punti costituiscono l’insieme dei punti più vicini $Y_k = \bigcup_{j=1}^N p_j$.
3. Si calcola la rototraslazione Q che minimizza la distanza media tra i punti trasformati $Q(x_j)$ e i p_j . Sia d_k il valore di tale distanza.
4. Si applica la rototraslazione: $X_{k+1} = Q(X_k)$
5. Se la differenza $d_k - d_{k-1}$ è inferiore ad una certa soglia τ , si arresta l’algoritmo, altrimenti si pone $k = k + 1$ e si torna al punto 2.

3.2 Formulazione in termini di densità di probabilità

Obiettivo della sezione è fornire una formulazione probabilistica delle mappe di distanza introdotte nella sezione **3.1.1**, che saranno oggetto del controllo statistico di processo di cui ci occuperemo nei capitoli 4 e 5. Per semplicità, siano X e Y due superfici bidimensionali chiuse e limitate (cosa che include il caso delle mesh a facce triangolari, ovviamente), opportunamente allineate come specificato nella sezione **3.1.3**. Vorremmo semplificare lo studio delle mappe d_X e d_Y . Rinunciamo allora all’indagine diretta della distribuzione dei valori di distanza nello spazio, preferendo piuttosto lavorare in termini probabilistici: qual è la probabilità che un punto scelto secondo una distribuzione uniforme su X si trovi ad una certa distanza da Y , e analogamente

scambiando i ruoli di X e Y ? Possiamo facilmente formalizzare quest'idea: concentriamoci sulla mappa $d_Y : X \rightarrow \mathbb{R}^+$ che, come già osservato, associa ad ogni $x \in X$ il valore di distanza (euclidea) $\inf_{y \in Y} d(x, y)$. Dalle ipotesi fatte, il codominio $C \subset \mathbb{R}^+$, $C = d_Y(X)$ è limitato e, se X e Y sono sufficientemente regolari, è un intervallo. Si definisca allora lo spazio di probabilità $(X, \mathcal{B}(X), \mathbb{P})$, dove \mathbb{P} è la misura di probabilità uniforme su X . É allora possibile definire la variabile aleatoria

$$D_Y : (X, \mathcal{B}(X)) \rightarrow (C, \mathcal{B}(C)) \text{ definita da } D_Y = d_Y(x)$$

In ipotesi di regolarità delle superfici, D_Y è assolutamente continua, quindi possiamo supporre l'esistenza di una densità f_{D_Y} per la legge \mathbb{P}^{D_Y} , di modo che

$$\mathbb{P}(a \leq D_Y \leq b) = \int_a^b f_{D_Y}(t) dt, \quad a, b \in C.$$

Analogamente, è immediato definire la variabile aleatoria

$$D_X : (Y, \mathcal{B}(Y)) \rightarrow (K, \mathcal{B}(K)) \text{ definita da } D_X = d_X(y)$$

dove K è un altro intervallo limitato. La definizione di f_{D_X} segue facilmente. Riassumendo, abbiamo ridotto le mappe

$$d_Y : X \rightarrow C, \quad d_X : Y \rightarrow K$$

alle funzioni di densità di probabilità

$$f_{D_Y} : C \rightarrow \mathbb{R}^+, \quad f_{D_X} : K \rightarrow \mathbb{R}^+,$$

passando quindi da funzioni reali aventi per dominio delle superfici a funzioni reali di variabile reale. Appare dunque evidente che occorre pagare un prezzo per questa semplificazione. In effetti, la semplice conoscenza delle funzioni di densità garantisce senz'altro informazioni sulla quantità di punti che si trovano a una certa distanza o, per essere più precisi, sulla probabilità che un punto estratto a caso su X si trovi a una certa distanza da Y . Tuttavia, la semplificazione ci porta a perdere tutte le informazioni spaziali, ovvero conoscere le funzioni di densità non ci permette di risalire alla parte dell'oggetto in cui i punti a distanza maggiore, e quindi i probabili difetti, si concentrano. Questo problema è per certi versi simile al problema della perdita di informazioni circa la correlazione spaziale nell'analisi di immagini fatto emergere in (Menafoglio et al. 2018). D'altro canto, si osservi che in un contesto di controllo di produzione questa semplificazione spaziale non è necessariamente negativa. Se l'obiettivo principale del controllo risulta essere l'identificazione di qualunque tipo di anomalia negli oggetti prodotti, indipendentemente da dove tale anomalia sia localizzata, risulterà infatti sufficiente studiare il profilo delle densità, che forniscono informazioni puntuali circa la quantità di massa dell'oggetto costituita da punti materiali che si trovano a una certa distanza dal prototipo.

3.2.1 Estensione ad un dataset di oggetti e proprietà della rappresentazione

Si cercherà adesso di calare la formulazione probabilistica appena introdotta in un contesto pratico, in modo da chiarire quali siano i passaggi che ci porteranno dal lavorare con un dataset di mesh ad un dataset di densità. Si supponga di disporre di un campione di mesh $\{S_i\}_{i=1,\dots,N}$, rappresentazioni di diverse realizzazioni dello stesso prototipo P , e siano tali mesh tutte allineate a P nello spazio come da sezione **3.1.3**. È possibile calcolare, in linea di principio, le N mappe

$$d_{S_i} : P \rightarrow C_i, \quad i = 1, \dots, N,$$

con C_i opportuni intervalli del semiasse positivo, e le N mappe

$$d_P^i : S_i \rightarrow K_i, \quad i = 1, \dots, N,$$

con K_i ancora intervalli. A questo punto, il significato di ciascuna mappa dovrebbe essere chiaro. Applicando quanto illustrato nella sezione precedente, possiamo far corrispondere ad ogni mappa una densità di probabilità, ottenendo quindi le $2N$ densità

$$f_{S_i} : C_i \rightarrow \mathbb{R}^+, \quad f_P^i : K_i \rightarrow \mathbb{R}^+.$$

Si hanno dunque a disposizione due dataset, il dataset \mathbf{f}_S contenente le f_{S_i} e il dataset \mathbf{f}_P contenente le f_P^i . Tali dataset corrispondono, come osservato alla fine della sezione **3.1.2**, a informazioni geometriche differenti. In riferimento al dataset concreto a nostra disposizione, ci aspettiamo quindi che gli elementi OOC-LS e OOC-HS, irregolari a causa di abbondanza di materiale sulle strutture trabecolari, possano essere classificati come difettosi solo guardando al dataset \mathbf{f}_P , mentre OOC-MS, caratterizzato da *assenza* di materiale, venga rilevato solo da un'analisi del dataset \mathbf{f}_S .

È il momento giusto per soffermarsi e riflettere sulle proprietà della rappresentazione che abbiamo costruito. Il dataset iniziale, costituito da N mesh a facce triangolari, è stato trasformato in due dataset, ognuno composto da N densità di probabilità che hanno per supporto degli intervalli di distanze euclidee. Riassumiamo di seguito le proprietà di tale soluzione, di cui alcune sono già state esplicitamente sottolineate, mentre altre verranno approfondite nel seguito:

- Come osservato nella sezione precedente, una notevole quantità di informazioni viene perduta nel processo di trasformazione, in quanto si tiene solo conto di come sono distribuite le distanze dei punti delle mesh rispetto al prototipo, e vice-versa. Ogni riferimento spaziale viene trascurato.

- Siamo in grado di interpretare prontamente, anche in maniera qualitativa, la rappresentazione che abbiamo scelto. Ad esempio, ad un picco di densità anomalo situato su distanze relativamente grandi si associa subito un'anomalia nella mesh associata alla densità.
- Si tratta di una rappresentazione *sintetica*, contenente solo informazioni necessarie alla soluzione del problema che stiamo attaccando, ovvero il monitoraggio e la quantificazione della gravità delle differenze degli oggetti prodotti rispetto al prototipo.
- Al fine di ottenere la rappresentazione dal dataset iniziale, è necessaria una fase di pre-processamento non banale e non facilmente automatizzabile, per via della necessità di allineamento che abbiamo trattato in sezione **3.1.3**

É chiaro, a questo punto, che il prossimo passo deve consistere nello sviluppo di metodi in grado di spiegare la variabilità all'interno di un dataset i cui atomi sono densità di probabilità. Questo sarà l'obiettivo dell'intero capitolo 4.

Capitolo 4

Spazi di densità di probabilità e SFPCA

4.1 Dati composizionali

Nel terzo capitolo, ed in particolare nella parte conclusiva, si è illustrato un metodo piuttosto generale per ridurre un dataset di N mesh, che abbiano un prototipo comune come riferimento, ai due dataset \mathbf{f}_S e \mathbf{f}_P , ognuno contenente N densità di probabilità. Si supponga, per semplicità, che gli elementi di \mathbf{f}_S abbiano tutte lo stesso supporto C , ovvero che $C_i = C \forall i = 1, \dots, N$ e analogamente che tutti gli elementi di \mathbf{f}_P abbiano lo stesso supporto K . Vedremo in seguito che tale semplificazione risulta ragionevole in un buon numero di situazioni pratiche. Concentriamoci su \mathbf{f}_P per fissare le idee, in quanto ogni considerazione che seguirà varrà per ambedue i dataset. Siano adesso $a, b \in K$, con $a < b$ due numeri reali appartenenti a K , che chiaramente ne individuano un sottointervallo $[a, b]$. Le N probabilità

$$\mathbb{P}(a \leq D_P^i \leq b) = \int_a^b f_P^i(t) dt$$

identificano, per ogni mesh, la *frazione* di punti della mesh stessa che si trovano a una distanza dal prototipo che sia compresa tra a e b . Pertanto, a profili di densità in qualche modo “anomali”, magari rispetto ad un insieme di elementi del dataset che si suppone essere “in controllo”, possono essere immediatamente associate anomale frazioni di massa ad un certo range di distanza dal prototipo, associabili a loro volta a difetti di produzione. L’analogia tra densità di probabilità e massa fisica è immediata per via del campionamento uniforme sulle mesh (sezione **3.2**). Occorre dunque monitorare il profilo di funzioni di densità, caratterizzate da due elementi critici:

- Si tratta di dati *funzionali*, ovvero ogni singolo elemento del dataset è infinito-dimensionale
- Si tratta di dati *composizionali*, cioè portatori di informazioni relative (frazioni dell'unità, appunto)

Il primo elemento caratterizzante richiede l'uso dei metodi della statistica per dati funzionali, per le cui basi si può ad esempio fare riferimento a (Ramsay e Silverman 2005). Negli ultimi decenni, i dati funzionali hanno costituito una sfida stimolante per gli statistici, portando allo sviluppo di teorie e tecniche adatte a trattare dati sempre più complessi (e.g. funzioni reali, funzioni a valori tensoriali, funzioni a valori su varietà riemanniane), volte ad estendere al dominio infinito-dimensionale le metodologie tipicamente applicate a dataset immersi in \mathbb{R}^n (test d'ipotesi, Principal Component Analysis, e così via). Limitandosi al caso delle funzioni reali, molte tecniche classiche possono essere estese senza troppe difficoltà teoriche, purchè il dataset di funzioni con cui si lavora possa essere immerso in un opportuno spazio di Hilbert separabile. Questo fatto è, almeno a posteriori, decisamente ragionevole, in quanto è ben noto che gli spazi di Hilbert separabili sono gli spazi a dimensione infinita più topologicamente simili a \mathbb{R}^n . La natura composizionale delle funzioni di densità, d'altro canto, richiede di unire l'analisi di dati a dimensione infinita alla teoria della *Compositional Data Analysis*. Tale campo della statistica nasce in seguito alla necessità di applicare i comuni metodi inferenziali, predittivi o di studio della variabilità a dataset multivariati i cui tipici elementi $\mathbf{x} \in \mathbb{R}^D$ sono vettori le cui componenti non possono prendere valori in tutto \mathbb{R}^D , ma nel simpleso (Filzmoser et al. 2018):

$$S^D = \{\mathbf{x} \in \mathbb{R}^D \text{ s. t. } x_i \geq 0 \forall i = 1, \dots, D, \sum_{i=1}^D x_i = \kappa\}$$

Dalla definizione si intuisce facilmente come il tipico oggetto di analisi siano vettori le cui componenti sono percentuali, probabilità o proporzioni. Appare altresì evidente, con un po' di immaginazione, come le funzioni di densità possano essere viste come una particolare generalizzazione infinito dimensionale degli elementi del simpleso, in quanto per definizione una funzione reale f è una densità se $f \geq 0$ e $\int f = 1$, vincoli analoghi a quelli di S^D . È noto da molto tempo, sin dalle osservazioni di Pearson, che applicare le usuali tecniche di inferenza a dati immersi nel simpleso, ignorandone la natura vincolata, porta a risultati errati, distorti o comunque di dubbia interpretabilità (Pearson 1897). Queste criticità sono in larga parte dovute al fatto che la maggior parte delle comuni tecniche di analisi si basa, più o meno implicitamente, sulla struttura vettoriale euclidea (Eaton 1983), o hilbertiana per il caso infinito dimensionale, dello spazio in cui il dataset assume i valori. Pertanto, trattare dati vincolati ad appartenere ad S^D come se fossero invece semplicemente elementi di \mathbb{R}^D è un procedimento scorretto,

anche solo per via del fatto che S^D non è uno spazio vettoriale rispetto alle usuali operazioni di somma e prodotto per uno scalare. Considerazioni perfettamente analoghe valgono nel caso in cui si trattino funzioni di densità a quadrato integrabile, ovvero gli elementi dell'insieme

$$A^2(a, b) = \{f : [a, b] \rightarrow \mathbb{R} \text{ t.c. } f \in L^2(a, b), f \geq 0, \int_a^b f(x) dx = 1\}$$

come elementi dello spazio $L^2(a, b)$, lo spazio di Hilbert più comune nello studio dei dati funzionali. Anche in questo caso, è immediato osservare che le operazioni di somma e prodotto per uno scalare tipiche di L^2 non generano su A^2 la struttura di spazio vettoriale. Le proprietà geometriche del simpleso sono state, negli ultimi decenni, ampiamente studiate, sin dal pionieristico lavoro (Aitchison 1982). Tale studio ha portato, in tempi più recenti, alla definizione di operazioni adeguate a rendere S^D uno spazio vettoriale con prodotto scalare, e la generalizzazione di tali operazioni ad opportuni spazi di densità costituisce la base di quella che si potrebbe definire *Compositional Functional Data Analysis*, di cui la SFPCA costituisce un'espressione. Obiettivo del capitolo è dunque fornire una panoramica molto sintetica della struttura geometrica del simpleso nel caso finito dimensionale, per poi mostrare come tale struttura possa essere estesa a spazi funzionali.

4.1.1 Geometria di Aitchison e operazioni nel simpleso

Chiamiamo composizione D -dimensionale un generico vettore $\mathbf{x} \in \mathbb{R}^D$ contenente solo informazioni relative. Chiamiamo subcomposizione di \mathbf{x} un vettore $\mathbf{x}^d \in \mathbb{R}^d$, $d < D$, contenente d componenti della composizione originale \mathbf{x} . È molto utile avere ben chiaro quali sono le proprietà fondamentali che i dati composizionali devono rispettare. Secondo (Egozcue 2009), e come ribadito in (Filzmoser et al. 2018), tali proprietà possono essere riassunte nei seguenti principi:

- **Invarianza per riscalamanti:** le informazioni contenute in un vettore di probabilità o percentuali non variano se tale vettore viene moltiplicato per una costante λ . Si tratta di un'osservazione molto facile da comprendere, in quanto è evidente che rappresentare percentuali in una scala da 1 a 100 o probabilità in una scala da 0 a 1 è una scelta totalmente arbitraria e convenzionale.
- **Invarianza per permutazioni:** le informazioni contenute in una composizione non variano se tale composizione subisce una permutazione delle sue componenti. Si tratta di un principio presente anche nella classica analisi statistica multivariata.
- **Coerenza per subcomposizioni:** in primo luogo, l'invarianza per riscalamanti *deve* essere valida anche per ogni sotto-vettore di una

composizione D -dimensionale \mathbf{x} . Inoltre le informazioni contenute in qualunque subcomposizione $\mathbf{x}^d \in \mathbb{R}^d$ di \mathbf{x} non possono essere in contraddizione con le informazioni contenute in \mathbf{x} . Questo fatto può essere espresso in maniera più formale. Sia infatti Δ_d una metrica definita sulle composizioni d -dimensionali, per un generico d . Allora deve valere

$$\Delta_d(\mathbf{x}^d, \mathbf{y}^d) \leq \Delta_D(\mathbf{x}, \mathbf{y}) \quad \forall \mathbf{x}, \mathbf{y}, d < D \quad (1)$$

Questa particolare proprietà prende il nome di dominanza per subcomposizioni. Si noti che, a prima vista, si tratta di una proprietà ragionevole anche per dati non composizionali, ed è certamente condivisa dalla metrica euclidea di \mathbb{R}^d . Vedremo tra poco, tuttavia, che quando accoppiata con la peculiare proprietà di invarianza per riscalamanti, la dominanza per subcomposizioni risulta avere conseguenze molto profonde.

Una conseguenza dei principi precedenti è che qualunque funzione di variabile composizionale, che esprima una qualche informazione significativa, può dipendere solo dai rapporti tra le parti (le componenti) della composizione, peculiarità che segue dalla proprietà di invarianza per riscalamanti. Tale proprietà, come già anticipato, e come trattato in maniera formale in (Barcelo-Vidal et al. 2001), risulta essere quella con le conseguenze più profonde, essendo quella che, di fatto, rende inadeguate le comuni tecniche di analisi. Se torniamo infatti alla relazione (1), e aggiungiamo la proprietà di invarianza, troviamo che deve essere:

$$\Delta_d(\mathbf{x}^d, \mathbf{y}^d) = \Delta_d(\lambda \mathbf{x}^d, \lambda \mathbf{y}^d) \leq \Delta_D(\lambda \mathbf{x}, \lambda \mathbf{y}) = \Delta_D(\mathbf{x}, \mathbf{y}) \quad \forall \mathbf{x}, \mathbf{y}, \lambda > 0, d < D$$

proprietà che è falsa ad esempio, per la metrica euclidea (e per tutte le p -metriche classiche in \mathbb{R}^d , ovvero quelle indotte dalle norme $\|\mathbf{x}\|^p := \sum_{i=1}^D x_i^p$, $p \geq 1$). Limitiamoci, senza perdita di generalità, al caso $\kappa = 1$, ovvero al simpleso standard

$$S^D = \{\mathbf{x} \in \mathbb{R}^D \text{ s. t. } x_i > 0 \quad \forall i = 1, \dots, D, \sum_{i=1}^D x_i = 1\}$$

e mostriamo ora le definizioni di somma e prodotto proposte da Aitchison per rendere S^D uno spazio vettoriale. Si noti che dal simpleso standard sono escluse le composizioni che presentano componenti nulle. L'operazione di somma tra \mathbf{x} e \mathbf{y} , chiamata perturbazione tra composizioni, è definita come

$$\mathbf{x} \oplus \mathbf{y} = \frac{1}{\sum_{i=1}^D x_i y_i} (x_1 y_1, \dots, x_D y_D)^T$$

che è evidentemente una somma ben definita in S^D . È interessante notare come tale definizione possa essere interpretata come “somma di informazioni”,

in senso Bayesiano. In effetti, se si interpretano \mathbf{x} e \mathbf{y} come leggi di variabili aleatorie discrete indipendenti X e Y a valori in $\{1, \dots, D\}$, è evidente che $\mathbf{x} \oplus \mathbf{y}$ rappresenta la legge condizionata di X , o equivalentemente quella di Y , dato l'evento $X = Y$. Si noti inoltre che, in perfetto accordo alle osservazioni precedenti, la definizione di somma è invariante rispetto a cambiamenti di scala delle composizioni.

L'operazione di prodotto per uno scalare si definisce come

$$\alpha \odot \mathbf{x} = \frac{1}{\sum_{i=1}^D x_i^\alpha} (x_1^\alpha, \dots, x_D^\alpha)^T$$

che è un prodotto ben definito per S^D sul campo scalare \mathbb{R} . Le due definizioni ci consentono di definire anche la differenza, come $\mathbf{x} \ominus \mathbf{y} = \mathbf{x} \oplus (-1 \odot \mathbf{y})$. Si può dare alla definizione un'interpretazione simile alla precedente. Se si suppone infatti che α sia un intero positivo, e che le variabili aleatorie $\{X_i\}_{i=1, \dots, \alpha}$ siano i.i.d con legge \mathbf{x} , possiamo interpretare $\alpha \odot \mathbf{x}$ come la legge di X_1 condizionata all'evento $X_1 = X_2 = \dots = X_\alpha$. Ancora più significativa risulta l'osservazione che, dalle definizioni precedenti, l'elemento neutro rispetto alla somma risulta essere

$$\mathbf{0} = (1/D, \dots, 1/D)^T$$

ovvero la distribuzione uniforme su $\{1, \dots, D\}$, facilmente interpretabile come la distribuzione "a minimo contenuto di informazioni", o anche la distribuzione con la massima entropia. Pertanto, assegnata \mathbf{x} come legge di X , si può interpretare $\ominus \mathbf{x}$ come la legge della variabile aleatoria Y che rende uniforme la legge condizionata $\mathbb{P}(X|X = Y)$.

In conclusione, la tripletta (S^D, \oplus, \odot) costituisce uno spazio vettoriale reale. Per generare su di esso una struttura geometrica coerente con le proprietà dei dati composizionali fatte emergere in precedenza, definiamo il *prodotto scalare di Aitchison* tra elementi di S^D come:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \frac{1}{2D} \sum_{i=1}^D \sum_{j=1}^D \log \frac{x_i}{x_j} \log \frac{y_i}{y_j}.$$

Tale prodotto scalare genera una norma e una metrica su S^D , chiamate *norma di Aitchison* e *distanza di Aitchison* rispettivamente:

$$\|\mathbf{x}\| = \sqrt{\frac{1}{2D} \sum_{i=1}^D \sum_{j=1}^D \left(\log \frac{x_i}{x_j}\right)^2}$$

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} \ominus \mathbf{y}\|.$$

Ci concentriamo adesso su un importante risultato riguardante il semplice, la sua dimensione e la sua completezza (Aitchison 1986). La dimostrazione di tale risultato è propedeutica alla comprensione di analoghi risultati in un contesto infinito dimensionale, e preferiamo dunque riportarla.

Teorema 4.1.1. *Il simpleso standard $S^D = \{\mathbf{x} \in \mathbb{R}^D \text{ s. t. } x_i > 0 \forall i = 1, \dots, D, \sum_{i=1}^D x_i = 1\}$, con le operazioni \oplus e \odot , munito del prodotto scalare di Aitchison, è linearmente isometricamente isomorfo al sottospazio di \mathbb{R}^D , munito dell'usuale prodotto scalare euclideo, costituito dai vettori la cui somma delle componenti è nulla. In particolare, S^D è uno spazio vettoriale con dimensione $D - 1$.*

Dimostrazione. Iniziamo col proporre l'isomorfismo, che indichiamo con $clr : S^D \rightarrow \mathbb{R}^D$, da "centered log ratio". Tale funzione è definita come segue: se $\mathbf{y} = clr(\mathbf{x})$, allora

$$y_i = \log x_i - \frac{1}{D} \sum_{j=1}^D \log x_j.$$

Tale funzione è continua nella topologia indotta dalla norma di Aitchison. Si tratta di una funzione iniettiva: supponiamo infatti che $\mathbf{x}_1 \neq \mathbf{x}_2$ e $clr(\mathbf{x}_1) = clr(\mathbf{x}_2)$. Quindi

$$\log x_{1i} - \frac{1}{D} \sum_{j=1}^D \log x_{1j} = \log x_{2i} - \frac{1}{D} \sum_{j=1}^D \log x_{2j}.$$

Prendendo l'esponenziale di ambo i membri si trova

$$\frac{x_{1i}}{(\prod_{j=1}^D x_{1j})^{\frac{1}{D}}} = \frac{x_{2i}}{(\prod_{j=1}^D x_{2j})^{\frac{1}{D}}}.$$

Sommando membro a membro e sfruttando il fatto che le composizioni hanno somma unitaria, troviamo immediatamente $\prod_{j=1}^D x_{1j} = \prod_{j=1}^D x_{2j}$ da cui $x_{1i} = x_{2i} \forall i$, da cui discende l'iniettività. Si vede che effettivamente \mathbf{y} è a somma nulla, in quanto

$$\sum_{i=1}^D y_i = \sum_{i=1}^D \log x_i - \frac{D}{D} \sum_{j=1}^D \log x_j = 0.$$

Questo mostra che $clr(S^D) \subset F^D$, dove con F^D abbiamo indicato i vettori a somma nulla. Dobbiamo mostrare che $F^D \subset clr(S^D)$ per mostrare che il candidato isomorfismo è suriettivo. Sia allora $\mathbf{y} \in F^D$, e definiamo la composizione $\mathbf{x} = clr^{-1}(\mathbf{y})$ come

$$x_i = \frac{e^{y_i}}{\sum_{j=1}^D e^{y_j}}.$$

Si vede immediatamente che \mathbf{x} è una composizione e che $clr(\mathbf{x}) = \mathbf{y}$. Quest'ultimo fatto si dimostra sfruttando $\sum_{j=1}^D y_j = 0$. Inoltre clr^{-1} è continuo. Proviamo che clr è lineare. Sia $\mathbf{y} = clr(\alpha \odot \mathbf{x}_1 \oplus \mathbf{x}_2)$. Abbiamo

$$(\alpha \odot \mathbf{x}_1 \oplus \mathbf{x}_2)_i = \frac{x_{1i}^\alpha x_{2i}}{\sum_{j=1}^D x_{1j}^\alpha x_{2j}}$$

pertanto

$$y_i = \alpha \log x_{1i} + \log x_{2i} - \log \sum_{j=1}^D x_{1j}^\alpha x_{2j} - \frac{1}{D} \sum_{j=1}^D (\alpha \log x_{1j} + \log x_{2j}) + \log \sum_{j=1}^D x_{1j}^\alpha x_{2j}$$

da cui si trova subito

$$\mathbf{y} = \alpha \text{clr}(\mathbf{x}_1) + \text{clr}(\mathbf{x}_2)$$

ovvero la linearità. Pertanto clr è una bigezione continua, con inversa continua, da S^D a F^D , ovvero clr è un isomorfismo. Per mostrare che è anche un'isometria, occorre verificare che

$$\langle \mathbf{y}_1, \mathbf{y}_2 \rangle = \langle \text{clr}^{-1}(\mathbf{y}_1), \text{clr}^{-1}(\mathbf{y}_2) \rangle$$

dove ovviamente il prodotto scalare a sinistra è quello euclideo, mentre quello a destra è il prodotto scalare di Aitchison. Per quanto possa sembrare poco elegante, è sufficiente sviluppare il membro di destra. Risulta infatti:

$$\langle \text{clr}^{-1}(\mathbf{y}_1), \text{clr}^{-1}(\mathbf{y}_2) \rangle = \frac{1}{2D} \sum_{i=1}^D \sum_{j=1}^D (y_{1i} - y_{1j})(y_{2i} - y_{2j}) =$$

$$\frac{1}{2D} (2D \langle \mathbf{y}_1, \mathbf{y}_2 \rangle) = \langle \mathbf{y}_1, \mathbf{y}_2 \rangle$$

dove nell'ultimo passaggio abbiamo sfruttato il fatto che \mathbf{y}_1 e \mathbf{y}_2 sono a somma nulla. \square

L'aver ottenuto un isomorfismo isometrico con un sottospazio di R^D risulta estremamente significativo, anche perché ciò ci permette di affermare che S^D è uno spazio metrico completo rispetto alla metrica indotta dal prodotto scalare di Aitchison. Questo potrebbe sembrare contro intuitivo, se si è abituati a ragionare in termini di metriche euclidee, in quanto presa ad esempio $\mathbf{x}_n = (1 - \frac{1}{n}, \frac{1}{n})$ si potrebbe pensare che \mathbf{x}_n non converge in S^2 perché $(1, 0) \notin S^2$. Questa obiezione, tuttavia, non è significativa, perché \mathbf{x}_n non è una successione di Cauchy in S^2 . Infatti per m ed n grandi

$$\|\mathbf{x}_n \ominus \mathbf{x}_m\|^2 \approx \frac{1}{2D} (2 \log^2 m + 2 \log^2 n - 4 \log m \log n)$$

che non converge a 0. Alla completezza di S^D si aggiunge anche un'importante utilità pratica dell'isomorfismo isometrico clr . Data infatti una base per F^D , applicando l'isomorfismo inverso si ottiene una base per S^D . È dunque possibile, ad esempio, trasformare un dataset di dati composizionali via clr , eseguire la Principal Component Analysis in F^D (costruendo una base di autovettori adeguata), per poi tornare indietro attraverso clr^{-1} . In tal modo, si è in grado di rispettare la struttura geometrica di S^D senza tuttavia dover lavorare con le peculiari operazioni di somma e prodotto definite in tale

spazio. Questo procedimento sarà alla base della SFPCA, ovvero analisi in componenti principali per funzioni di densità, che adatteremo in questo lavoro. Procederemo adesso col mostrare una generalizzazione, piuttosto recente, della geometria del simpleso ad opportuni spazi di densità, dopo una brevissima parentesi sugli spazi di Hilbert.

4.2 Spazi di Hilbert

Come illustrato nella sezione precedente, per poter applicare tecniche di riduzione dimensionale come la Principal Component Analysis, di analisi di varianza in stile ANOVA, e molti altri metodi usuali dell'analisi statistica multivariata, è necessario che gli elementi del dataset possano essere pensati come elementi di uno spazio, anche infinito dimensionale, in cui possano essere definiti i concetti di ortogonalità, proiezione e base ortonormale. In altre parole, è necessario che lo spazio a cui si fa riferimento abbia la struttura di spazio di Hilbert separabile. Ripercorriamo ora, molto brevemente, le principali definizioni e i risultati fondamentali derivati dallo studio di tali spazi. Nel passare in rassegna tali risultati, si daranno per note le nozioni di spazio vettoriale, norma, spazio normato, completezza, spazio di Banach, spazio duale. Si tenga presente che l'obiettivo della sezione non è una presentazione completa ed esaustiva della teoria degli spazi di Hilbert, che sarebbe del tutto al di fuori degli scopi di questo lavoro. Si desidera invece affrontare un breve percorso ragionato che illustri come e perchè le proprietà di questi spazi sono spesso strumenti estremamente utili nell'analisi di dati funzionali o comunque complessi. Il lettore interessato ad approfondire in maniera sistematica questa ampia branca dell'analisi funzionale può fare riferimento, ad esempio, a (Rudin 1991).

4.2.1 Prime definizioni e proprietà principali

La seguente definizione è ben nota, in quanto di fondamentale importanza in moltissimi campi della matematica e della fisica. Si tratta della nozione di *prodotto scalare*, che costituisce il fondamento dell'intera teoria degli spazi di Hilbert.

Definizione 4.2.1. *Sia H uno spazio vettoriale reale. Un prodotto scalare è una forma bilineare, simmetrica e definita positiva su H , ovvero una funzione*

$$\langle \cdot, \cdot \rangle : H \times H \rightarrow \mathbb{R}$$

che soddisfa le seguenti proprietà:

1. $\langle \alpha u + v, w \rangle = \alpha \langle u, w \rangle + \langle v, w \rangle \forall \alpha \in \mathbb{R}, \forall u, v, w \in H$ (*proprietà di linearità*)
2. $\langle u, v \rangle = \langle v, u \rangle \forall u, v \in H$ (*proprietà di simmetria*)

3. $\langle u, u \rangle \geq 0 \forall u \in H$ (proprietà di positività)
4. $\langle u, u \rangle = 0$ se e solo se $u = 0$ (proprietà di annullamento)

In perfetto accordo con quanto si fa con i vettori di \mathbb{R}^n , diciamo che due elementi u e v di H sono tra loro ortogonali se $\langle u, v \rangle = 0$. La seguente proposizione riassume le proprietà più elementari del prodotto scalare in spazi vettoriali reali.

Proposizione 4.2.1. *Sia H uno spazio vettoriale reale dotato di prodotto scalare. Definiamo su H la norma indotta dal prodotto scalare come $\|u\| = \sqrt{\langle u, u \rangle} \forall u \in H$. Valgono allora i seguenti risultati*

1. $\langle u, v \rangle \leq \|u\| \|v\|$ (Disuguaglianza di Cauchy-Shwartz)
2. La coppia $(H, \|\cdot\|)$ costituisce uno spazio vettoriale normato.
3. Il prodotto scalare è una forma bilineare continua in ambedue gli argomenti
4. $\|u + v\|^2 = \|u\|^2 + \|v\|^2$ se e solo se u e v sono tra loro ortogonali (Teorema di Pitagora)
5. Se $\sum_{n=1}^{\infty} u_n$ è una serie convergente in H , e $\{u_n\}_{n \geq 1}$ è un insieme di elementi di H a due a due ortogonali, allora $\|\sum_{n=1}^{\infty} u_n\|^2 = \sum_{n=1}^{\infty} \|u_n\|^2$

Siamo adesso in grado di introdurre la definizione cruciale, quella di spazio di Hilbert.

Definizione 4.2.2. *Uno spazio vettoriale H , dotato di prodotto scalare, si dice spazio di Hilbert se la coppia $(H, \|\cdot\|)$, con $\|\cdot\|$ indotta dal prodotto scalare, è uno spazio di Banach, ovvero se $(H, \|\cdot\|)$ è uno spazio vettoriale normato e completo.*

Pertanto, tutti gli spazi di Hilbert sono spazi di Banach, ma il viceversa ovviamente non vale (si pensi allo spazio di Banach L^1 delle funzioni sommabili). La proprietà di completezza risulta cruciale nella dimostrazione delle proprietà proiettive degli spazi di Hilbert, proprietà che, come già osservato, sono spesso alla base della Statistica su dati infinito-dimensionali. Si noti che, facendo riferimento alla sezione precedente, S^2 è uno spazio di Hilbert con il prodotto scalare di Aitchison. Enunciamo adesso i teoremi che ci consentono di poter effettuare proiezioni, e quindi analisi in componenti principali, negli spazi di Hilbert.

Teorema 4.2.1. *Sia H uno spazio di Hilbert e $C \subset H$ un suo sottoinsieme chiuso, convesso e non vuoto. Sia $u \in H \setminus C$. Allora esiste uno e un solo $w \in C$ tale che*

$$d(u, w) = \inf_{v \in C} d(v, w)$$

dove la distanza si intende indotta dalla norma.

Se V è un sottospazio chiuso di H , è evidentemente convesso, e l'elemento w prende il nome di proiezione ortogonale di u su V . Dato un sottospazio chiuso V , chiamiamo allora V^\perp l'insieme

$$V^\perp := \{u \in H \text{ t.c. } \langle u, v \rangle = 0 \ \forall v \in V\}.$$

Tale insieme si chiama sottospazio ortogonale a V , ed è sempre chiuso. Vale il seguente fondamentale risultato, noto come Teorema delle Proiezioni.

Teorema 4.2.2. *Sia H uno spazio di Hilbert e V un suo sottospazio chiuso. Allora*

$$H = V \oplus V^\perp$$

ovvero ogni elemento w di H è rappresentabile in modo unico come somma $w = u + v$ con $u \in V$ e $v \in V^\perp$.

Il vettore u viene generalmente chiamato *proiezione ortogonale* di w su V . Questo teorema, insieme a quelli che seguono, risulta fondamentale nella dimostrazione dell'analogo infinito dimensionale del teorema spettrale per matrici simmetriche definite positive, risultato che è alla base della comune PCA per dataset standard.

Gli spazi di Hilbert hanno anche la peculiare caratteristica di essere *sempre* rappresentabili attraverso *basi ortonormali*:

Definizione 4.2.3. *Sia H uno spazio di Hilbert. $A \subset H$ è una base ortonormale per H se soddisfa le seguenti proprietà:*

- $\|u\| = 1 \ \forall u \in A$
- Se $u_1, u_2 \in A$, $u_1 \neq u_2$, allora $\langle u_1, u_2 \rangle = 0$
- $w \in H$ è tale che $\langle w, u \rangle = 0 \ \forall u \in A$ se e solo se $w = 0$

Abbiamo allora il seguente risultato di esistenza per le basi ortonormali

Teorema 4.2.3. *Ogni spazio di Hilbert ammette una base ortonormale. Ogni spazio di Hilbert separabile ammette una base ortonormale numerabile.*

Si osservi che l'esistenza di una base numerabile *non* è garantita nel caso degli spazi di Banach, anche se riflessivi. Questa è un'ulteriore caratteristica che rende gli spazi di Hilbert separabili gli spazi a dimensione infinita "più vicini" al comune spazio euclideo. In effetti, vale per le basi ortonormali un risultato del tutto analogo al caso delle basi ortonormali in \mathbb{R}^n .

Proposizione 4.2.2. *Sia H uno spazio di Hilbert e $\{u_n\}_{n \geq 0}$ una sua base ortonormale. Allora, $\forall x \in H$:*

1. $x = \sum_{n=0}^{\infty} \langle x, u_n \rangle u_n$ (Sviluppo in serie di Fourier)

2. $\langle x, y \rangle = \sum_{n=0}^{\infty} \langle x, u_n \rangle \langle y, u_n \rangle$ (Uguaglianza di Parseval)
3. $\|x\|^2 = \sum_{n=0}^{\infty} \langle x, u_n \rangle^2$

Enunciamo infine, come anticipato in precedenza, il Teorema Spettrale in spazi di Hilbert. Supporremo note le nozioni di compattezza e simmetria di un operatore lineare, così come anche quelle di autovalori e autovettori.

Teorema 4.2.4. *Sia $K : H \rightarrow H$ un operatore lineare, compatto e simmetrico, H uno spazio di Hilbert separabile. Allora i suoi autovalori distinti formano una successione decrescente e tendente a 0. Inoltre gli autovettori di K possono essere scelti in modo da formare una base ortonormale di H .*

In particolare, un operatore compatto e simmetrico K può essere rappresentato come segue:

$$Ku = \sum_{j=1}^{\infty} \lambda_j \langle u, w_j \rangle w_j$$

Dove i λ_j sono gli autovalori di K e i w_j i corrispondenti autovettori.

4.2.2 Osservazioni e PCA in spazi di funzioni

I risultati rapidamente passati in rassegna nella sezione precedente rendono possibile generalizzare la PCA ad ogni spazio funzionale che abbia la struttura di spazio di Hilbert separabile. Per mostrare il procedimento in pratica, ci restringeremo senza perdere troppa generalità al caso dello spazio

$$L^2(\Omega) = \{f : \Omega \rightarrow \mathbb{R} \text{ misurabile t.c. } \int_{\Omega} f^2(t) dt < +\infty\}$$

col prodotto scalare $\langle f, g \rangle = \int_{\Omega} f(t)g(t) dt$, seguendo il procedimento di (Horváth e Kokoszka 2012). Supponiamo allora che $\{X_i\}_{i=1, \dots, N}$ sia un campione funzionale aleatorio estratto da $L^2(\Omega)$, che senza perdita di generalità supporremo centrato rispetto alla media campionaria $\frac{1}{N} \sum_{i=1}^N X_i$, calcolata ovviamente con le usuali operazioni di somma e prodotto per uno scalare definite in L^2 . La prima componente principale del campione, volendo estendere il procedimento classico, è l'elemento $w_j \in L^2(\Omega)$ che risolve il problema di massimizzazione

$$\max_{u \in L^2(\Omega)} \sum_{j=1}^N \langle u, X_j \rangle^2$$

con il vincolo $\|u\| = 1$. Le altre componenti risolvono problemi analoghi, con in più vincoli di ortogonalità rispetto alle componenti principali trovate in precedenza. Si può dimostrare che (ibid.) le componenti principali sono esattamente gli autovettori dell'operatore di covarianza campionario, definito come $C : L^2(\Omega) \rightarrow L^2(\Omega)$ tale che

$$Cv = \frac{1}{N} \sum_{j=1}^N \langle X_j, v \rangle X_j$$

Tali autovettori si intendono ordinati in rispetto all'ordinamento decrescente dei rispettivi autovalori, garantito dal Teorema Spettrale enunciato nella sezione precedente. In maniera analoga al caso finito dimensionale, l'autovalore corrispondente ad una componente principale può essere interpretato come la “quantità di varianza” spiegata dalla componente. Tale operatore, avente rango finito, risulta compatto, ed è chiaramente simmetrico. Pertanto esso soddisfa le ipotesi del Teorema Spettrale e ammette la decomposizione vista alla fine della sezione precedente, il che implica che le componenti principali sono ben definite. Diversi metodi sono disponibili per la risoluzione esplicita del problema (Ramsay e Silverman 2005), tra cui ad esempio la rappresentazione degli elementi del campione attraverso i primi K elementi di una base ortonormale di L^2 , che riduce dunque il problema ad una PCA multivariata classica.

Il prossimo passo consiste nella definizione di un opportuno spazio di Hilbert di densità di probabilità, che renda dunque possibile manipolare via PCA il dataset introdotto nel capitolo 1. Tale passo, in definitiva, costituisce l'unione tra analisi di dati funzionali e di dati composizionali.

4.3 Lo spazio B^2 e le sue caratteristiche

Seguendo un procedimento simile a quello di (Pawlowsky-Glahn et al. 2014), procediamo generalizzando le proprietà del simpleso evidenziate nella sezione 4.1.1 ad uno spazio di Hilbert di densità. Come vedremo, il processo sarà leggermente più complicato del caso finito dimensionale.

Sia (Ω, \mathcal{A}) uno spazio misurabile. Sia λ una misura finita definita su tale spazio. Si consideri lo spazio

$$B(\lambda) := \{\mu \text{ t.c. } \mu \text{ misura su } (\Omega, \mathcal{A}), \mu \text{ a.c. rispetto a } \lambda, \lambda \text{ a.c. rispetto a } \mu\}$$

dove per a.c. si intende “assolutamente continua”. In termini più semplici, stiamo definendo lo spazio delle misure equivalenti a λ . Si noti che al momento gli elementi di $B(\lambda)$ non soddisfano alcun vincolo analogo a quello del simpleso. Introduciamo allora la seguente relazione di equivalenza:

Definizione 4.3.1. Diciamo che $\mu = \nu$ in $B(\lambda)$ se e solo se esiste $c > 0$ tale che

$$\mu(A) = c\nu(A) \quad \forall A \in \mathcal{A},$$

dove si intende $c \cdot \infty = \infty$

Dalla definizione segue che tutte le misure finite di $B(\lambda)$ hanno un rappresentante che è una misura di probabilità. L'unica stranezza rispetto al simpleso standard, pertanto, è costituita dal fatto che in $B(\lambda)$ sono presenti misure infinite. Si tratta di una complicazione necessaria per ottenere spazi di densità completi (di Banach e di Hilbert).

4.3.1 Somma e prodotto

Il teorema di Radon-Nikodym garantisce che ogni elemento di $B(\lambda)$ è rappresentabile dalla propria derivata rispetto alla misura di riferimento λ . Ciò ci consente di generalizzare in maniera immediata le operazioni che rendono S^D uno spazio vettoriale.

Definizione 4.3.2. *Siano $\mu, \nu \in B(\lambda), \alpha \in \mathbb{R}$. Definiamo in $B(\lambda)$ la somma tra μ e ν come segue:*

$$\mu \oplus \nu(A) := \int_A \frac{d\mu}{d\lambda} \frac{d\nu}{d\lambda} d\lambda \quad \forall A \in \mathcal{A}$$

Definiamo il prodotto per uno scalare in $B(\lambda)$ come segue:

$$\alpha \odot \mu(A) := \int_A \frac{d\mu^\alpha}{d\lambda} d\lambda$$

Le definizioni sono date a meno della relazione di equivalenza definita nella sezione precedente.

Si verifica facilmente che tali operazioni sono effettivamente una somma e un prodotto ben definiti e che rendono $B(\lambda)$ uno spazio vettoriale. D'ora in poi, per rendere la notazione più leggera e più vicina agli obiettivi pratici, useremo il teorema di Radon-Nikodym per identificare ogni misura con la propria derivata rispetto a λ , identificando dunque $B(\lambda)$ con lo spazio $\{f : \Omega \rightarrow \mathbb{R} \text{ misurabile t.c. } f > 0 \text{ q.o.}\}$, che continueremo a chiamare $B(\lambda)$. Pertanto la definizione precedente viene modificata come segue:

$$f \oplus g := fg$$

$$\alpha \odot f := f^\alpha$$

Tali formule risultano più compatte e più simili alle formule analoghe nel simpleso. Le definizioni sono in questo caso date a meno di una nuova relazione di equivalenza tra le funzioni di $B(\lambda)$, ovvero

$$f_1 = f_2 \text{ se e solo se } \exists c > 0 \text{ t.c. } f_1 = cf_2 \text{ quasi ovunque.}$$

Si noti che la formula del prodotto è ben data in quanto $f > 0$ quasi ovunque, e che, in analogia al caso delle misure, ogni elemento di $B(\lambda)$ che sia integrabile ha per rappresentante una densità di probabilità. Osserviamo che le formule precedenti ci consentono di definire anche la differenza, che riportiamo per comodità:

$$f \ominus g = \frac{f}{g}.$$

Procediamo adesso alla definizione degli spazi $B^p(\lambda)$. In tutto il seguito, con $B(\lambda)$ si intenderà sinteticamente lo spazio vettoriale reale $(B(\lambda), \oplus, \odot)$.

4.3.2 Spazi di Banach di densità

Con l'intento di definire una classe di spazi di Banach di densità con una norma analoga a quella di Aitchison, procediamo definendo per ogni $p \geq 1$ i sottospazi vettoriali di $B(\lambda)$:

$$B^p(\lambda) := \{f \in B(\lambda) \text{ t.c. } \log f \in L^p(\Omega, \lambda)\}.$$

Si potrebbe essere tentati dal normare $B^p(\lambda)$ in maniera ovvia, con $\|f\|_{B^p} = \|\log f\|_{L^p}$. Tuttavia, una definizione del genere sarebbe priva di senso, in quanto dipendente dalla scelta del rappresentante delle classi di equivalenza della relazione definita in precedenza. Ciò è intimamente collegato al fatto che una simile scelta sarebbe in aperta contraddizione col principio di *invarianza per riscalamenti* in sezione 4.1.1. Pertanto, imitando la norma di Aitchison in S^D definiamo

$$\|f\|_{B^p}^p = \frac{1}{2\lambda(\Omega)} \iint \left| \log \frac{f(x)}{f(y)} \right|^p d\lambda(x) d\lambda(y),$$

dove gli integrali si intendono su tutto lo spazio. Si vede subito che la norma non dipende dal rappresentante scelto per calcolarla. Inoltre notiamo che la definizione è ben data:

$$\|f\|_{B^p}^p = \frac{1}{2\lambda(\Omega)} \iint |\log f(x) - \log f(y)|^p d\lambda(x) d\lambda(y) \leq \|\log f\|_{L^p}^p < \infty,$$

dove abbiamo usato la disuguaglianza triangolare per il valore assoluto nell'ultimo passaggio. Per quanto riguarda l'omogeneità:

$$\|\alpha \odot f\|_{B^p}^p = \frac{1}{2\lambda(\Omega)} \iint |\alpha|^p \left| \log \frac{f(x)}{f(y)} \right|^p d\lambda(x) d\lambda(y) = |\alpha|^p \|f\|_{B^p}^p.$$

La disuguaglianza triangolare è parimenti immediata, anche se leggermente più laboriosa. Notiamo che

$$\|f\|_{B^p}^p = \frac{1}{2\lambda(\Omega)} \left\| \log \frac{f(x)}{f(y)} \right\|_{L^p(\lambda \otimes \lambda)}^p.$$

Abbiamo:

$$\begin{aligned} \|f \oplus g\|_{B^p}^p &= \frac{1}{2\lambda(\Omega)} \iint |\log f(x) - \log f(y) + \log g(x) - \log g(y)|^p d\lambda(x) d\lambda(y) = \\ &= \frac{1}{2\lambda(\Omega)} \left\| \log \frac{f(x)}{f(y)} + \log \frac{g(x)}{g(y)} \right\|_{L^p(\lambda \otimes \lambda)}^p \leq \frac{1}{2\lambda(\Omega)} \left(\left\| \log \frac{f(x)}{f(y)} \right\|_{L^p(\lambda \otimes \lambda)} + \left\| \log \frac{g(x)}{g(y)} \right\|_{L^p(\lambda \otimes \lambda)} \right)^p \end{aligned}$$

ed estraendo la radice p -esima troviamo la disuguaglianza triangolare. La proprietà di annullamento è immediata. Infatti se la norma di f è nulla, allora $\log \frac{f(x)}{f(y)}$ è nullo quasi ovunque, per cui $f(x)$ è quasi ovunque uguale ad una

costante, pertanto f è lo 0 di $B(\lambda)$. Così come nel caso finito dimensionale, è opportuno costruire un isomorfismo che consenta di ricavare le proprietà topologiche di B^p da quelle di L^p senza troppi sforzi. Tenendo a mente le proprietà della funzione clr in S^D , definiamo

$$L_0^p(\lambda) := \{f \in L^p(\lambda) \text{ t.c. } \int_{\Omega} f d\lambda = 0\}$$

sottospazio di $L^p(\lambda)$ delle funzioni a media nulla. Si osservi che la definizione è ben data in quanto $\int_{\Omega} f$ è finito, poichè $L^p(\lambda) \subset L^1(\lambda)$ essendo λ finita. Che si tratti in effetti di un sottospazio è ovvio, tuttavia vale anche il seguente risultato, un semplice esercizio di analisi funzionale:

Proposizione 4.3.1. $L_0^p(\lambda)$ è chiuso in $L^p(\lambda)$ e di conseguenza è uno spazio di Banach con la stessa norma di $L^p(\lambda)$.

Dimostrazione. Sia f un punto di accumulazione per L_0^p . Quindi esiste una successione $f_n \in L_0^p$ convergente a f in L^p . Segue che f_n converge a f in L^1 perchè λ è finita. La convergenza in L^1 forza la convergenza della successione degli integrali, dunque $\int_{\Omega} f_n d\lambda \rightarrow \int_{\Omega} f d\lambda$. Segue immediatamente che $f \in L_0^p$. Quindi L_0^p è chiuso perchè contiene i suoi punti di accumulazione, perciò è anche completo, ovvero è di Banach. \square

Siamo pronti a definire l'isomorfismo clr nel nuovo contesto:

Teorema 4.3.1. Sia $clr : B^p(\lambda) \rightarrow L_0^p(\lambda)$ definito come segue:

$$clr(f) = \log f - \frac{1}{\lambda(\Omega)} \int_{\Omega} \log f d\lambda$$

Allora clr è lineare, continuo, invertibile, con inverso continuo.

Dimostrazione. Che clr sia iniettivo e lineare si dimostra esattamente come nella sezione 4.1.1 sostituendo gli integrali alle sommatorie, così come la relazione $clr(B^p(\lambda)) \subset L_0^p(\lambda)$. Sia $h \in L_0^p(\lambda)$, si definisca $clr^{-1}(h) = e^h$. Allora $clr^{-1}(h)$ è evidentemente in $B(\lambda)$. Inoltre

$$clr(clr^{-1}(h)) = h + \frac{1}{\lambda(\Omega)} \int_{\Omega} h d\lambda = h$$

perchè h è a media nulla. Dunque clr è suriettivo, per cui invertibile con inversa clr^{-1} . È immediato notare che anche clr^{-1} è lineare. Passiamo alla continuità:

$$clr(f)(x) = \log f(x) - \frac{1}{\lambda(\Omega)} \int_{\Omega} \log f d\lambda = \frac{1}{\lambda(\Omega)} \int_{\Omega} \log \frac{f(x)}{f(y)} d\lambda(y)$$

da cui

$$\|clr(f)\|_{L^p}^p = \frac{1}{\lambda(\Omega)^p} \int_{\Omega} \left| \int_{\Omega} \log \frac{f(x)}{f(y)} d\lambda(y) \right|^p d\lambda(x) \leq$$

$$\leq \frac{2}{\lambda(\Omega)^{p-1}} \|f\|_{B^p}^p$$

dove si è usata la disuguaglianza di Jensen nell'ultimo passaggio. La relazione ottenuta risolve il problema della continuità. Resta solo da mostrare la continuità dell'inversa clr^{-1} :

$$\begin{aligned} \|clr^{-1}(h)\|_{B^p}^p &= \frac{1}{2\lambda(\Omega)} \iint |h(x) - h(y)|^p d\lambda(x) d\lambda(y) \leq \\ &\leq \frac{1}{2\lambda(\Omega)} \iint (|h(x)|^p + |h(y)|^p) d\lambda(x) d\lambda(y) = \|h\|_{L^p}^p \end{aligned}$$

che conclude la dimostrazione. \square

A questo punto, il prossimo passo è ovvio:

Proposizione 4.3.2. *Lo spazio vettoriale normato $(B^p(\lambda), \|\cdot\|_{B^p})$ è uno spazio di Banach separabile.*

Dimostrazione. Sia $\{f_n\}_{n \geq 1}$ una successione fondamentale (di Cauchy) in $B^p(\lambda)$. Per definizione, allora:

$$\|f_n \ominus f_m\|_{B^p} \rightarrow 0 \text{ se } m, n \rightarrow \infty.$$

Sia $\{h_n\}_{n \geq 1} \subset L_0^p(\lambda)$ definita da

$$h_n = clr(f_n) \forall n \geq 1.$$

Poichè clr è continuo e lineare $\{h_n\}_{n \geq 1}$ è una successione fondamentale di $L_0^p(\lambda)$. Ma $L_0^p(\lambda)$ è uno spazio di Banach, per cui h_n converge a un limite h in $L_0^p(\lambda)$. Sia allora

$$f = clr^{-1}(h).$$

Abbiamo dunque, per via della continuità di clr^{-1} :

$$\|f_n - f\|_{B^p} = \|clr^{-1}(h_n - h)\|_{B^p} \leq C \|h_n - h\|_{L^p}.$$

Il membro di destra tende a 0 per $n \rightarrow \infty$, forzando a 0 anche il membro a sinistra. Quindi $\{f_n\}_{n \geq 1}$ converge a f in $B^p(\lambda)$. In conclusione, tutte le successioni fondamentali di $B^p(\lambda)$ sono convergenti, il che rende $B^p(\lambda)$ uno spazio di Banach. La separabilità segue in maniera analoga, sfruttando le proprietà di clr e del suo inverso. \square

Enunciamo adesso dei risultati riguardo alcuni sottoinsiemi di $B^p(\lambda)$, dimostrati nel caso $p = 2$ in (Pawlowsky-Glahn et al. 2014), e immediatamente generalizzabili per $p \geq 1$, in quanto non dipendenti dalla struttura hilbertiana. Premettiamo:

Definizione 4.3.3. *Definiamo il sottospazio di $B(\lambda)$ delle densità doppiamente limitate:*

$$B_b(\lambda) = \{f \in B(\lambda) \text{ s.t. } \exists c > 0 \text{ s.t. } \frac{1}{c} \leq f \leq c\}$$

Il sottoinsieme delle densità di probabilità:

$$B_P(\lambda) = \{f \in B(\lambda) \text{ s.t. } f \in L^1(\lambda)\}$$

Il sottoinsieme delle densità non integrabili:

$$B_I(\lambda) = B(\lambda) \setminus B_P(\lambda)$$

e in maniera ovvia definiamo

$$B_b^p(\lambda) = B_b(\lambda) \cap B^p(\lambda), \quad B_P^p(\lambda) = B_P(\lambda) \cap B^p(\lambda), \quad B_I^p(\lambda) = B_I(\lambda) \cap B^p(\lambda)$$

Grazie a (Boogaart et al. 2010), è noto che B_b è un sottospazio, mentre B_P è convesso, quindi lo stesso vale per B_b^p e B_P^p . È ovvio, inoltre, che B_P è contenuto in B_b . Il risultato più interessante è il seguente:

Proposizione 4.3.3. *Il sottoinsieme $B_b^p(\lambda)$ di $B^p(\lambda)$ è denso in $B^p(\lambda)$. Di conseguenza, lo è anche $B_P^p(\lambda)$.*

Dimostrazione. L'immagine di $B_b^p(\lambda)$ in $L_0^p(\lambda)$ attraverso clr è costituita dalle funzioni limitate di $L_0^p(\lambda)$. Tali funzioni sono dense in $L_0^p(\lambda) = clr(B^p(\lambda))$ quindi $B_b^p(\lambda)$ è denso in $B^p(\lambda)$. \square

Concludiamo la sezione con alcune osservazioni:

- La definizione dello spazio vettoriale $B(\lambda)$, delle relazioni di equivalenza e degli spazi $B^p(\lambda)$ è interamente contenuta nel lavoro di (Pawlowsky-Glahn et al. 2014), così come la definizione e alcune delle proprietà dell'operatore clr .
- Tuttavia, tale lavoro è principalmente concentrato sulla struttura di spazio di Hilbert di $B^2(\lambda)$, utile nelle applicazioni pratiche, e pertanto non vengono esplicitamente sfruttate le proprietà di clr per conferire una struttura di spazio di Banach a $B^p(\lambda)$. In particolare, $B^p(\lambda)$ non viene normato e le proprietà di completezza di $B^2(\lambda)$ vengono dimostrate direttamente.
- È sembrato più naturale, in questo lavoro, normare $B^p(\lambda)$ con una norma analoga a quella indotta dal prodotto scalare di S^D , e dedurre le proprietà di tali spazi per poi ricavare naturalmente, come faremo nella prossima sezione, la struttura Hilbertiana di $B^2(\lambda)$ mediante la definizione di un prodotto scalare che ne induca la norma.

- Sembra che in nessuno lavoro attualmente presente in letteratura $B^p(\lambda)$ venga normato e di conseguenza dotato di una struttura di spazio di Banach.
- La densità dello spazio delle densità di probabilità nel più ampio spazio B^p costituirà un elemento fondamentale ai fini della piena comprensione dello spazio di Hilbert B^2 : approfondiremo questo aspetto nella prossima sezione

4.3.3 Lo spazio B^2

Definiamo adesso su $B^2(\lambda)$ la seguente forma bilineare (Egozcue et al. 2006), ovviamente ispirata al prodotto scalare di Aitchison:

$$\langle f, g \rangle_{B^2} = \frac{1}{2\lambda(\Omega)} \iint \log \frac{f(x)}{f(y)} \log \frac{g(x)}{g(y)} d\lambda(x) d\lambda(y)$$

Possiamo allora ricavare immediatamente il seguente risultato:

Teorema 4.3.2. *Vale la seguente relazione:*

$$\langle \cdot, \cdot \rangle_{B^2} = \langle clr(\cdot), clr(\cdot) \rangle_{L^2}.$$

Tale relazione ha le seguenti conseguenze:

1. $\langle \cdot, \cdot \rangle_{B^2}$ è un prodotto scalare, perchè clr è un isomorfismo lineare e $\langle \cdot, \cdot \rangle_{L^2}$ è un prodotto scalare.
2. $B^2(\lambda)$ munito di tale prodotto scalare è uno spazio di Hilbert separabile, perchè è di Banach, è separabile, e la sua norma è indotta dal prodotto scalare.
3. $clr : B^2(\lambda) \rightarrow L_0^2(\lambda)$ oltre ad essere un isomorfismo è anche un'isometria.

Dimostrazione. È chiaro che dobbiamo solo dimostrare la relazione tra i prodotti interni, in quanto le conseguenze seguono immediatamente. La dimostrazione, basata su semplici calcoli, segue un procedimento analogo a (ibid.). Gli integrali si intendono tutti estesi all'intero spazio. Procediamo ponendo $\Lambda = \lambda(\Omega)$, $dx = d\lambda(x)$ per brevità:

$$\langle clr(f), clr(g) \rangle_{L^2} = \int \left\{ \left(\log f(x) - \frac{1}{\Lambda} \int \log f(y) dy \right) \left(\log g(x) - \frac{1}{\Lambda} \int \log g(y) dy \right) \right\} dx.$$

Sviluppando il prodotto e notando che vale l'ovvia relazione

$$\int \log f(x) dx \int \log g(y) dy = \int \log f(y) dy \int \log g(x) dx$$

si ottiene

$$\langle clr(f), clr(g) \rangle_{L^2} = \int \log f(x) \log g(x) dx - \frac{1}{\Lambda} \int \log f(x) dx \int \log g(y) dy.$$

Il secondo membro è uguale a

$$\frac{1}{\Lambda} \iint \left(\log f(x) \log g(x) - \log f(y) \log g(y) \right) dx dy$$

che può anche essere scritto come

$$\frac{1}{2\Lambda} \iint \left(\log f(x) \log g(x) + \log f(y) \log g(y) - \log f(x) \log g(y) - \log f(y) \log g(x) \right) dx dy.$$

Sfruttando le proprietà elementari dei logaritmi applicate a quest'ultima relazione concludiamo che

$$\langle clr(f), clr(g) \rangle_{L^2} = \frac{1}{2\lambda(\Omega)} \iint \log \frac{f(x)}{f(y)} \log \frac{g(x)}{g(y)} d\lambda(x) d\lambda(y) = \langle f, g \rangle_{B^2}.$$

□

Come si può notare, e come era stato anticipato alla fine della sezione precedente, la struttura Hilbertiana di B^2 è un'ovvia conseguenza della sua struttura di spazio di Banach, una volta dimostrata l'isometria oggetto del teorema appena introdotto. Occorre adesso comprendere qual è il ruolo dell'insieme delle densità integrabili, ovvero delle vere e proprie densità di probabilità, all'interno del più vasto spazio B^2 , così come è necessario costruire basi ortonormali di B^2 . Il risultato che segue è dovuto a (Egozcue et al. 2006), ed è accompagnato da una dimostrazione più breve, in quanto basata sui risultati ottenuti in precedenza

Teorema 4.3.3. *Sia $\{\psi_j\}_{j \geq 1}$ una base ortonormale di $L_0^2(\lambda)$ formata da funzioni limitate. Si definisca $\phi_j = clr^{-1}(\psi_j)$. Allora $\{\phi_j\}_{j \geq 1}$ è una successione in $B_b^2(\lambda)$, ed è una base ortonormale di $B^2(\lambda)$.*

Dimostrazione. Le ψ_j sono limitate, dunque $clr^{-1}(\psi_j) = e^{\psi_j}$ risultano doppiamente limitate e pertanto in $L^1(\lambda)$ dato che λ è finita. Segue che $\{\phi_j\}_{j \geq 1} \subset B_b^2(\lambda)$. Inoltre, $\{\phi_j\}_{j \geq 1}$ è ortonormale perchè $\{\psi_j\}_{j \geq 1}$ è ortonormale e clr è un isomorfismo isometrico. Infine, osserviamo che l'unico elemento u di $L_0^2(\lambda)$ tale che $\langle \psi_j, u \rangle_{L^2} = 0 \forall j$ è per definizione lo zero di $L_0^2(\lambda)$. Quindi, data l'isometria, l'unico elemento w di $B^2(\lambda)$ tale che $\langle \phi_j, w \rangle_{B^2} = 0 \forall j$ risulta essere $clr^{-1}(0)$ ossia lo 0 di $B^2(\lambda)$ e quindi $\{\phi_j\}_{j \geq 1}$ è una base ortonormale di $B^2(\lambda)$. □

Dal teorema precedente e dalla proposizione 4.2.2 sulle serie di Fourier, si evince che ogni elemento di B^2 , comprese le densità improprie, può essere scritto come somma infinita (somma \oplus , nel senso di $B(\lambda)$) di densità di probabilità doppiamente limitate, un fatto coerente con la densità di $B_b^p(\lambda)$ in $B^p(\lambda)$. Inoltre, sappiamo che qualunque combinazione lineare *finita* di elementi di $B_b^p(\lambda)$ genera a sua volta una densità doppiamente limitata, perchè $B_b^p(\lambda)$ è un sottospazio. Questa caratteristica si rivela estremamente utile, in quanto ogni volta che si costruisce una base di $B^2(\lambda)$ a partire da una base *limitata* di $L_0^2(\lambda)$, è possibile approssimare un campione di densità di probabilità usando i primi K elementi della base (come da sezione 4.2.2), avendo la garanzia che il sottospazio finito dimensionale con cui si finisce per lavorare contiene *solamente* densità di probabilità, senza mai includere funzioni non integrabili.

4.4 PCA nello spazio B^2

Avendo dotato B^2 della struttura di spazio di Hilbert separabile, siamo adesso pronti ad illustrare in dettaglio l'analisi delle componenti principali per un dataset di densità di probabilità. Seguiremo l'approccio di (Hron et al. 2014), in cui viene proposta una versione della PCA per dati funzionali (FPCA) adattata alla geometria dello spazio B^2 , e che dipende fortemente dalle proprietà dell'isomorfismo *clr*, ampiamente trattato nella sezione precedente. Tale versione della FPCA prende il nome di Simplicial Functional Principal Component Analysis (SFPCA). Per semplicità, ci restringeremo al caso in cui $\Omega = I$ con I intervallo limitato e λ è la misura di Lebesgue ristretta a I . In altre parole, ci stiamo concentrando sullo spazio $B^2(I)$ delle densità (proprie o improprie) che abbiano logaritmo in $L^2(I)$. In maniera simile a quanto fatto nella sezione 4.2.2, sia $\{X_i\}_{i=1\dots N}$ un dataset di densità di $B^2(I)$, centrato rispetto alla media campionaria funzionale $\bar{X} = \frac{1}{N} \odot \bigoplus_{i=1}^N X_i$. Per tutto il seguito, anche se non strettamente necessario dal punto di vista teorico, supporremo che gli elementi del dataset siano tutte densità integrabili, e che di conseguenza esse possano essere rappresentate da una densità di probabilità attraverso la relazione di equivalenza di $B^2(I)$. Definiamo allora la prima componente principale del dataset come

$$\zeta_1 = \arg \max_{\substack{u \in B^2(I) \\ \|u\|_{B^2} = 1}} \frac{1}{N} \sum_{i=1}^N \langle X_i, u \rangle_{B^2}^2$$

mentre le componenti ζ_j , $j = 2\dots N$ sono definite in maniera analoga, con l'aggiunta dei vincoli di ortogonalità $\langle \zeta_j, \zeta_k \rangle_{B^2} = 0 \ \forall k < j$. Esattamente come nel caso di L^2 (Horváth e Kokoszka 2012), la j -esima componente è il j -esimo autovettore dell'operatore di covarianza $V : B^2(I) \rightarrow B^2(I)$ la cui

azione è definita da

$$Vu = \frac{1}{N} \odot \bigoplus_{i=1}^N \langle X_j, u \rangle_{B^2} X_j \quad \forall u \in B^2(I).$$

Pertanto ogni coppia (λ_j, ζ_j) autovalore-componente soddisfa l'equazione agli autovalori

$$V\zeta_j = \lambda_j \odot \zeta_j.$$

Poichè clr è un'isomorfismo, possiamo applicarlo a entrambi i membri e sfruttarne la linearità trovando

$$\frac{1}{N} \sum_{i=1}^N \langle X_j, \zeta_j \rangle_{B^2} clr(X_j) = \lambda_j clr(\zeta_j).$$

Sfruttando l'isometria, troviamo infine

$$\frac{1}{N} \sum_{i=1}^N \langle clr(X_j), clr(\zeta_j) \rangle_{L^2} clr(X_j) = \lambda_j clr(\zeta_j).$$

Dall'ultima relazione, si ricava immediatamente che le coppie autovalore-autovettore dell'operatore di covarianza V sono in corrispondenza biunivoca con le coppie autovalore-autovettore (ρ_j, ξ_j) dell'operatore di covarianza $V_{clr} : L_0^2(I) \rightarrow L_0^2(I)$ la cui azione è definita da

$$V_{clr}x = \frac{1}{N} \sum_{i=1}^N \langle clr(X_j), x \rangle_{L^2} clr(X_j) \quad \forall x \in L_0^2(I).$$

In particolare, (λ_j, ζ_j) è la j -esima coppia autovalore-autovettore di V se e solo se (ρ_j, ξ_j) , con $\rho_j = \lambda_j$ e $\xi_j = clr(\zeta_j)$ è la j -esima coppia autovalore-autovettore per V_{clr} . Scegliamo adesso, per procedere al calcolo pratico degli autovalori, una base ortonormale *limitata* di $L_0^2(I)$ ed un intero K che tronchi tale base, in modo da beneficiare delle proprietà di sottospazio di $B_b^2(I)$, come osservato alla fine della sezione **4.3**. Ad esempio, seguendo il procedimento di (Machalová et al. 2015), si potrebbe scegliere una base di funzioni spline che soddisfi la condizione di integrale nullo. In ogni caso, indipendentemente dalla scelta della base, che chiamiamo $\{\phi_j\}_{j \geq 1}$, possiamo scrivere, approssimando L_0^2 col sottospazio finito-dimensionale generato dai primi K elementi della base:

$$clr(X_i) = \sum_{k=1}^K c_{ik} \phi_k$$

$$\xi_j = \sum_{k=1}^K b_{jk} \phi_k$$

in cui supponiamo che la matrice $C = c_{ik}$ sia nota, magari ottenuta attraverso fitting ai minimi quadrati sugli elementi del dataset trasformato, mentre gli N vettori matrice $\mathbf{b}_j = b_{jk}$ sono incogniti. Con questa notazione, possiamo approssimare la j -esima equazione agli autovalori

$$V_{clr}\xi_j = \lambda_j\xi_j, j = 1, \dots, K$$

con il sistema lineare

$$\frac{1}{N}C^T C\mathbf{b}_j = \rho_j\mathbf{b}_j.$$

A questo punto, le coppie $(\lambda_j, \mathbf{b}_j)$ possono essere calcolate attraverso PCA standard sulla matrice C . Ottenute tali coppie, possiamo immediatamente calcolare le componenti principali del dataset originale applicando l'isomorfismo inverso:

$$\zeta_j = \bigoplus_{k=1}^K b_{jk} \odot clr^{-1}(\phi_k),$$

nonchè la matrice di trasformazione $Z = CB^{-1}$, detta matrice degli *scores*, tale che

$$X_i = \bigoplus_{k=1}^K z_{ik} \odot \zeta_k.$$

Osserviamo che la scelta esplicita di una base non è l'unica possibilità di risolvere il problema della SFPCA. In particolare, come fatto nel nostro caso studio, è possibile usare degli stimatori delle funzioni di densità basati su kernel o su polinomi di Bernstein (Menafoglio et al. 2018) per poi discretizzare opportunamente il dominio I . In ogni caso, il calcolo degli autovalori λ_j , delle componenti ζ_j e della matrice Z risolve completamente il problema pratico della SFPCA, lasciandoci tuttavia l'onere di interpretare i risultati ottenuti, fase raramente banale e fortemente dipendente dalle caratteristiche dello specifico problema applicativo. Come vedremo nel corso del capitolo 5, diversi strumenti interpretativi sviluppati da (Ramsay e Silverman 2005), e.g., i grafici della media campionaria perturbata rispetto alle componenti principali, si riveleranno molto potenti quando accoppiati con il significato geometrico delle mappe di distanza trattate nel capitolo 3.

4.4.1 Possibili carte di controllo derivanti da SFPCA e conclusioni

Lasciamo per il momento da parte il problema interpretativo, e concentriamoci sul ricavare degli strumenti di Controllo Statistico di Processo a partire dai risultati ottenuti dalla SFPCA su un certo dataset centrato di densità di probabilità $\{X_i\}_{i=1\dots N} \subset B^2(I)$. Come già osservato, tali risultati sono riassunti dai K autovalori λ_j , dalle componenti ζ_j e dalla matrice degli scores Z . In (Menafoglio et al. 2018), vengono proposte due carte di controllo basate su tali elementi, e nello stesso lavoro viene mostrata l'effettiva superiorità

di tali carte rispetto a strumenti di controllo basati su FPCA classica, che ignorano dunque la geometria di Aitchison. La prima carta di controllo è basata su una generalizzazione della classica carta di controllo T^2 , basata sulla statistica T^2 di Hotelling. Occorre compiere un passo preliminare, ovvero scegliere quante delle K componenti effettivamente calcolate (al più $N - 1$) si desidera considerare nell'analisi. La conoscenza degli autovalori λ_j ci consente di applicare i criteri classici di scelta (Johnson e Wichern 2007):

- Fissata una soglia di varianza spiegata τ , si seleziona il minimo numero B di componenti tale che la frazione di varianza spiegata sia maggiore di τ :

$$B = \min \left\{ r \text{ t.c. } \frac{\sum_{j=1}^r \lambda_j}{\sum_{i=1}^K \lambda_i} \geq \tau \right\}$$

- Si sceglie B in corrispondenza del “gomito” nel grafico della varianza spiegata cumulata.

Un interessante elemento fatto emergere da (Menafoglio et al. 2018) consiste nell'osservazione che, anche nel caso di dataset di densità simulate e non attinenti a processi reali, il numero B di componenti principali necessarie a spiegare una certa percentuale di varianza risulta sensibilmente maggiore nel caso in cui si applichi alle densità una semplice FPCA ambientata in L^2 , non tenente conto della natura composizionale dei dati. Si tratta di un'ulteriore conferma del fatto che i dataset formati da dati composizionali, di qualunque dimensionalità, devono essere trattati con tecniche che tengano debito conto delle proprietà geometriche di questi dati. Tornando alla carta T^2 , scelto il numero di componenti, definiamo la statistica T^2 come segue:

$$T_j^2 = \sum_{k=1}^r \frac{z_{jk}^2}{\lambda_j}, \quad j = 1 \dots N$$

dove z_{jk} è lo score del j -esimo elemento del dataset, lungo la componente k -esima. Tale statistica, come si può facilmente immaginare, è analoga alla statistica T^2 di Hotelling nel caso multivariato, ed è utile a rilevare se un dato elemento del dataset si comporta in maniera anomala rispetto alla media, quando proiettato sugli spazi generati dalle componenti principali. Si tratta di un indicatore scalare, sintetico, e si presta dunque ad essere usato per la costruzione di una carta di controllo. Essendo per definizione un indicatore positivo, è sufficiente definire il limite di controllo superiore (UCL), che può essere determinato attraverso il calcolo di quantili empirici, ovvero l'opzione con meno difficoltà teoriche, o con delle ipotesi distribuzionali sul campione $\{T_j^2\}_{j=1, \dots, N}$. La seconda carta proposta in (ibid.) consiste nel monitoraggio dell'errore quadratico di ricostruzione degli elementi del dataset rispetto alle prime B componenti principali scelte, ovvero nel calcolo delle statistiche

$$Q_j = \left\| X_j - \bigoplus_{k=1}^B z_{jk} \odot \zeta_k \right\|_{B^2}^2$$

In questo caso, le statistiche Q sono utili a determinare se sono presenti elementi del dataset che vengono ricostruiti dalle prime B componenti con un'approssimazione più grossolana rispetto alla media. La carta T^2 e, in misura minore, la carta Q costituiranno, nel prossimo capitolo, gli esempi principali per mostrare come determinati difetti di produzione, anche in oggetti molto complessi, possano essere rilevati da un meccanismo di controllo che sia fondato da una parte su un solido e interpretabile significato geometrico (le mappe di distanza di cui al capitolo 3), e dall'altra su una rigorosa teoria matematica, basata sulla piena comprensione dei dati con cui si lavora (dati funzionali e composizionali).

Capitolo 5

Caso studio - Prima parte

5.1 Fase preliminare

Nel corso del capitolo 3 sono stati delineati i passi teorici necessari per ridurre un dataset di mesh ai due dataset di densità \mathbf{f}_P e \mathbf{f}_S . Ricordiamo che \mathbf{f}_P è il dataset contenente le densità relative alle distanze dei punti degli oggetti reali dal prototipo, mentre per \mathbf{f}_S vale il discorso inverso. In questa sezione, tali passi verranno studiati in maniera più approfondita, e si cercherà di spiegare in dettaglio tutta la fase di pre-processing delle mesh. Non si tratta di una fase particolarmente lunga o complicata, in quanto gli strumenti software di manipolazione di mesh e calcolo delle distanze a disposizione sono molti e ben consolidati, come osservato alla fine del capitolo 2. In tutto il capitolo, faremo riferimento al dataset materialmente a nostra disposizione, ma molte considerazioni possono essere estese senza difficoltà a qualunque dataset di mesh.

5.1.1 Codifica, pulizia e allineamento delle mesh

L'output della tomografia degli N elementi del campione consiste, ovviamente, in N mesh. Ogni mesh è definita dalla *Point Cloud* dei suoi vertici, e dalle facce triangolari che la compongono. Pertanto ciascuna mesh j , con $j = 1, \dots, N$, può essere facilmente codificata in un file di testo, in cui sono presenti:

- Una lista di n_j triplette di numeri reali $\mathbf{x}_k, k = 1, \dots, n_j$. La k -esima tripletta corrisponde alle coordinate cartesiane del k -esimo vertice della mesh.
- Una lista di m_j triplette di numeri interi $\mathbf{z}_r, r = 1, \dots, m_j$. In particolare, la r -esima tripletta contiene gli indici corrispondenti ai vertici della r -esima faccia della mesh.

Le due liste precedenti contengono tutte le informazioni geometriche relative alla mesh, in quanto da esse è possibile dedurre, in maniera diretta o indiretta,

la localizzazione spaziale di vertici, spigoli e facce della mesh, pertanto esse sono sufficienti per effettuare tutte le operazioni di manipolazione dei poliedri implementate nella libreria VCG, di cui abbiamo accennato nel capitolo 2. In particolare, i seguenti passi preliminari sono fondamentali per poter procedere al calcolo delle distanze:

1. È necessaria una fase di pulizia preliminare delle mesh, in quanto generalmente i rilevamenti effettuati dai comuni tomografi tendono a presentare artefatti e costrutti geometrici di piccole dimensioni che non corrispondono alla realtà, e che devono essere eliminati prima di analizzare la mesh. In particolare è necessario rimuovere tutti i vertici che non risultano appartenere ad alcuna faccia, nonché tutti gli spigoli che non abbiano esattamente due facce insistenti. Inoltre è necessario eliminare tutte le componenti della mesh che risultino sconnesse dalla componente più grande. Tutte queste operazioni, dato il file di codifica della mesh, possono essere svolte in maniera totalmente automatica, ad esempio attraverso l'interfaccia R di VCG, ossia la libreria `rvcg` menzionata nel capitolo 2.
2. Dopo la pulizia, è necessario che tutte le mesh corrispondenti agli oggetti prodotti vengano spazialmente allineate al prototipo nominale. Come già osservato, si può usare il consolidato algoritmo ICP in questa fase, che è però difficilmente automatizzabile, come osservato in dettaglio nella sezione **3.1.3**.

Avendo a disposizione delle mesh pulite e allineate, siamo pronti al calcolo delle mappe di distanza.

5.1.2 Calcolo delle mappe, stima delle densità e riduzione a un supporto unico

Facendo riferimento alla notazione definita nella sezione **3.2.1**, è necessario conoscere le $2N$ mappe

$$d_{S_j} : P \rightarrow C_j, \quad j = 1, \dots, N,$$

$$d_P^j : S_j \rightarrow K_j, \quad j = 1, \dots, N.$$

Non è ragionevole pensare di poter risalire ad un'espressione analitica per tali mappe, che pertanto vengono sempre calcolate approssimando le mesh con un opportuno sottoinsieme di punti: si rende dunque necessario discretizzare le mesh S_i e P , approssimandole con *Point Clouds*. È possibile applicare diversi criteri di discretizzazione (Cignoni et al. 1998), ad esempio è possibile limitarsi ad approssimare le mesh con le *Point Clouds* costituite dai rispettivi vertici, o anche procedere ad un campionamento uniforme di punti dalle facce. Nel seguito faremo riferimento al primo metodo, quindi i punti che

verranno considerati nel calcolo delle mappe saranno tutti e soli i vertici delle mesh. Non si tratta di una scelta vincolante, in quanto il criterio con cui si discretizzano i poliedri risulterà del tutto indipendente dalle analisi successive.

A questo punto, ogni mappa d_{S_j} consisterà in un vettore di n_P distanze, il cui k -esimo elemento corrisponde alla distanza del k -esimo vertice del prototipo P dal punto di S_j più vicino. Analogamente, ogni mappa d_P^j consiste di un vettore di n_j distanze, il cui k -esimo elemento corrisponde alla distanza del k -esimo vertice della mesh S_j dal punto di P più vicino. In termini probabilistici, per ogni mesh S_j abbiamo a disposizione un campione di n_P distanze (la mappa d_{S_j}) e un campione di n_j distanze (la mappa d_P^j): tali campioni sono utili a stimare le corrispondenti densità di probabilità. Focalizziamoci, per brevità, solo sulla mappa d_P^j , che pensiamo adesso come un campione aleatorio, i cui elementi verranno indicati con $d_k, k = 1, \dots, n_j$. È necessario stimare la densità $f_P^j : C_j \rightarrow \mathbb{R}^+$ a partire da tale campione, o da una trasformazione dello stesso (nel nostro caso, abbiamo adottato una trasformazione logaritmica). Qualunque metodo di stima, parametrico o non parametrico, è in linea di principio applicabile a tale problema, e non rientra negli scopi della tesi fornire indicazioni generali su quale sia il metodo “migliore”. Per questo caso studio, applicheremo un metodo di interpolazione basato sui polinomi di Bernstein, che si dimostra particolarmente adatto nel caso di densità a supporto compatto, e che è già stato applicato con successo in ambito di Controllo Statistico di Processo (Menafoglio et al. 2018). Sia $\bar{F}_P^j : C_j \rightarrow [0, 1]$ la funzione di ripartizione campionaria calcolata a partire dal campione d_P^j . Al fine di ottenere una funzione di ripartizione differenziabile, proponiamo lo stimatore (Babu et al. 2002):

$$\hat{F}_P^j(d) = \sum_{k=0}^{B_j} \bar{F}_P^j(k/B_j) b_{k, B_j}(d), \quad d \in C_j$$

in cui B_j è un parametro intero da scegliere in accordo alla numerosità n_j del campione mentre b_{k, B_j} è il k -esimo polinomio di Bernstein relativo a B_j . In (ibid.) viene dimostrata la consistenza di tale stimatore, da cui deriva lo stimatore per la funzione di densità f_P^j , nella forma

$$f_P^j(d) = B_j \sum_{k=0}^{B_j-1} (\bar{F}_P^j((k+1)/B_j) - \bar{F}_P^j(k/B_j)) b_{k, B_j-1}(d), \quad d \in C_j \quad (5.1)$$

Indipendentemente dalla scelta del metodo di stima delle densità, a valle del processo si ottengono i dataset \mathbf{f}_P e \mathbf{f}_S introdotti nel capitolo 2, in una forma “quasi adatta” alle analisi SFPCA. In effetti, prima di poter procedere all’analisi in componenti principali, è necessario tenere conto del fatto che tutte le densità hanno supporti diversi. Con riferimento al dataset \mathbf{f}_S , è necessario ridurre tutti i C_j ad un unico intervallo C , in modo da poter

lavorare nello spazio $B^2(C)$ come da sezione 4.4. È ragionevole immaginare, dato che le mesh sono tutte allineate, che le differenze tra i C_j siano minime e che, in particolare, esista un intervallo che sia contenuto in tutti i C_j . In questo caso studio si è pertanto deciso di scegliere il più grande intervallo C che sia contenuto in tutti i C_j , e di eliminare dai campioni d_{S_j} tutte le osservazioni che ricadano al di fuori di tale intervallo prima di procedere alla stima delle densità. Considerazioni analoghe valgono per il dataset \mathbf{f}_P , in cui tutti gli intervalli K_j devono essere ridotti ad un unico intervallo K , che può ovviamente essere diverso da C . Si osservi che non è possibile estendere le densità in maniera banale in modo da rendere il supporto il più grande possibile, prendendo l'unione dei C_j , in quanto le densità degli spazi B^p sono quasi ovunque strettamente positive. La riduzione del supporto costituisce l'ultimo passo preliminare necessario all'avvio della vera e propria fase di analisi, che procediamo ad illustrare.

5.2 Analisi del dataset \mathbf{f}_P

Questa sezione è interamente dedicata all'analisi del dataset \mathbf{f}_P , da cui ci aspettiamo, come già anticipato, l'individuazione degli elementi difettosi OOC-LS e OOC-HS, in quanto le mappe di distanza da cui ha origine il dataset contengono, tra le altre, le informazioni relative a difetti geometrici dovuti ad *abbondanza* di materiale. Si tenga presente che tutte le densità sono state stimate a partire dai campioni dei logaritmi delle distanze, per cui l'intervallo C a cui si fa riferimento nella sezione precedente *non* è vincolato ad appartenere al semiasse positivo, e nel seguito, con abuso di notazione, con d_P^j si intenderanno i dataset dei logaritmi delle distanze originali.

5.2.1 Ispezione del dataset e stima delle densità

Mostriamo nelle figure da 5.1 a 5.4 gli istogrammi relativi ai 16 campioni d_p^i . In ascissa è possibile leggere il nome identificativo degli elementi a cui si riferiscono gli istogrammi. Tutte le distanze a cui si fa riferimento, qui e nel seguito, sono in millimetri.

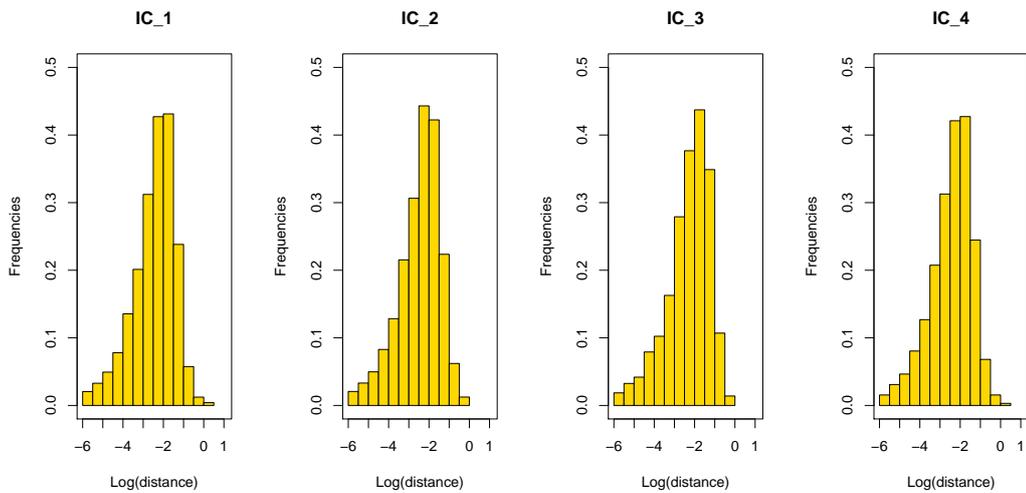


Fig.5.1: istogrammi delle frequenze relative per i logaritmi delle distanze, elementi IC1-IC4

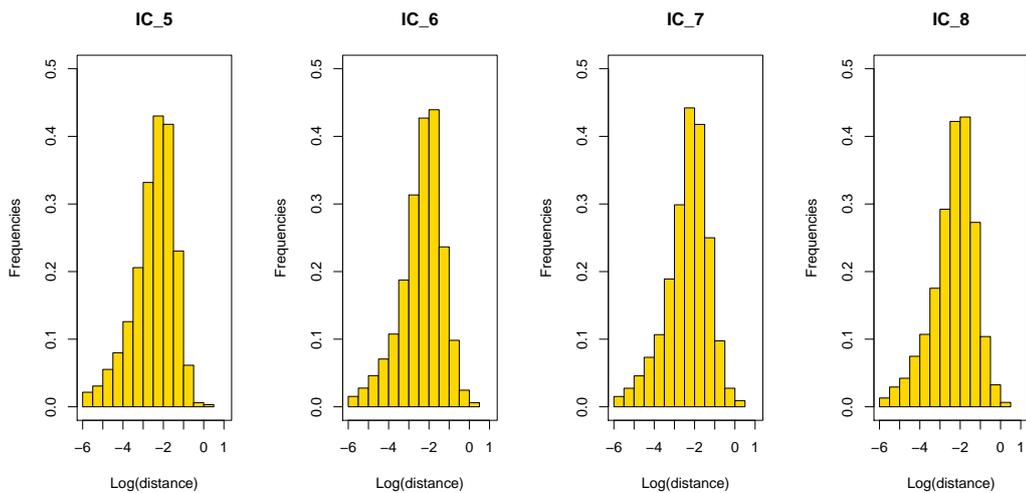


Fig.5.2: istogrammi delle frequenze relative per i logaritmi delle distanze, elementi IC5-IC8

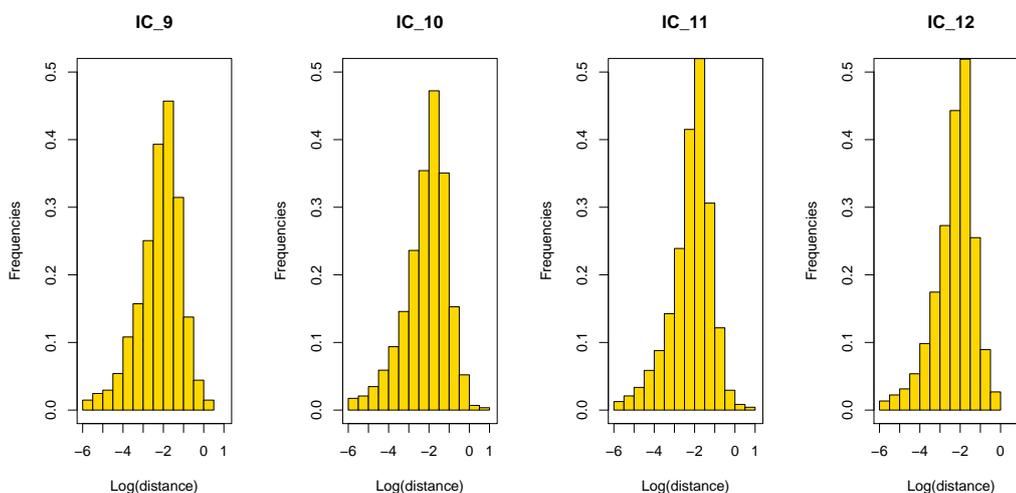


Fig.5.3: istogrammi delle frequenze relative per i logaritmi delle distanze, elementi IC9-IC12

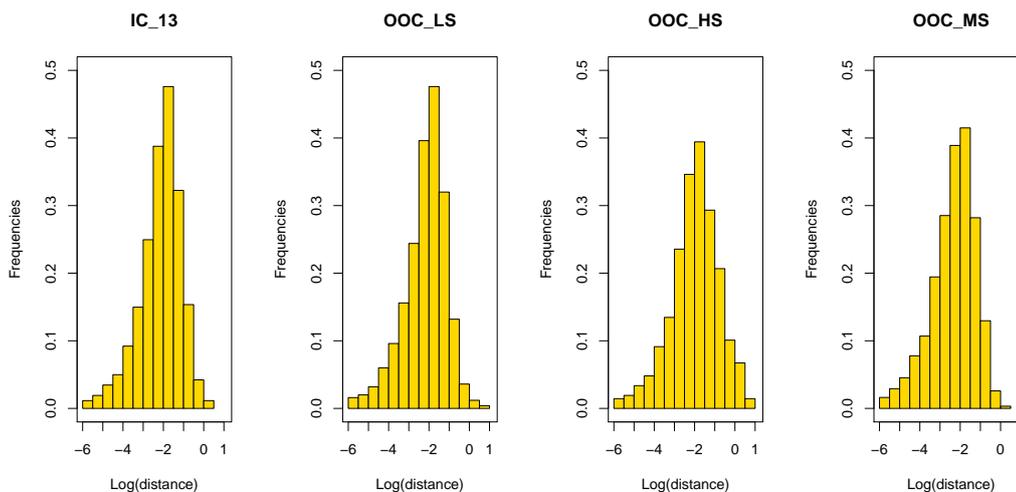


Fig.5.4: istogrammi delle frequenze relative per i logaritmi delle distanze, elementi IC13, OOC-LS, OOC-HS, OOC-MS

Si nota subito che le differenze da istogramma a istogramma non sono macroscopiche e, almeno da un'ispezione visiva, i profili degli istogrammi sono piuttosto simili. Si tratta di un fatto ragionevole, in quanto anche nel caso di elementi difettosi, le differenze da elemento a elemento non interessano parti delle mesh molto estese. In altre parole, la frazione di punti appartenenti a parti di mesh interessate da difetti risulta piccola rispetto al totale dei punti: abbiamo a che fare con difetti "locali". È facile immaginare, d'altro canto, che in caso di gravi difetti estesi all'intera mesh, e.g., struttura trabecolare del tutto deformata rispetto al prototipo, anche una semplice ispezione degli istogrammi classificherebbe immediatamente l'elemento come difettoso.

Procediamo adesso come mostrato nella sezione precedente, calcolando gli stimatori delle funzioni di ripartizione e di densità relativi ai precedenti istogrammi. In Fig.5.5 mostriamo un confronto tra le funzioni di ripartizione campionarie e e gli stimatori \hat{F} basati sui polinomi di Bernstein, introdotti nella precedente sezione. In questo caso, abbiamo scelto $B_j = n_j$, in accordo alle osservazioni di (Menafoglio et al. 2018).

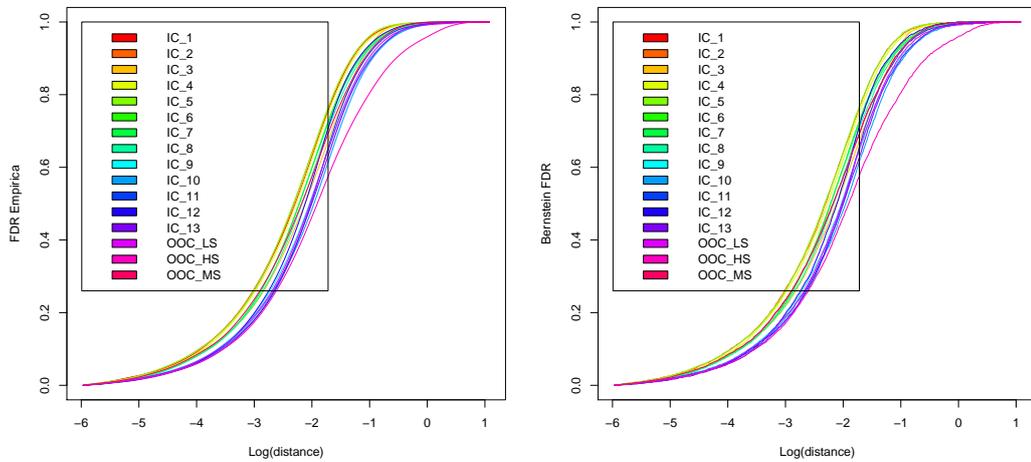


Fig.5.5: a sinistra le FDR empiriche, a destra gli stimatori \hat{F} . Si può apprezzare il fatto che gli stimatori lisci siano piuttosto simili alle FDR.

Si nota immediatamente che i profili delle funzioni di ripartizione stimate riproducono in maniera molto fedele i profili delle empiriche, col vantaggio addizionale di essere funzioni lisce, e quindi a densità continua. Inoltre, già a questo stadio, si nota un certo scostamento della funzione di ripartizione relativa a OOC-HS (in fucsia) rispetto al resto del campione. Questo fatto ci rassicura, poichè OOC-HS è l'elemento caratterizzato dai difetti più gravi, se ci restringiamo ai difetti rilevabili dall'analisi del dataset \mathbf{f}_P . Inoltre, l'andamento della funzione di ripartizione OOC-HS rispecchia fedelmente l'interpretazione geometrica delle distanze: il fatto che essa saturi a 1 più lentamente delle altre funzioni, è indice di una distribuzione di distanza con una coda più consistente su valori di distanza alti, il che è immediatamente associabile ad una frazione relativamente grande di punti della mesh interessati da difetti (si torni alla sezione **3.2.1** per maggiori dettagli). Mostriamo i corrispondenti stimatori delle funzioni di densità, calcolati come in (5.1), in Fig. 5.6.

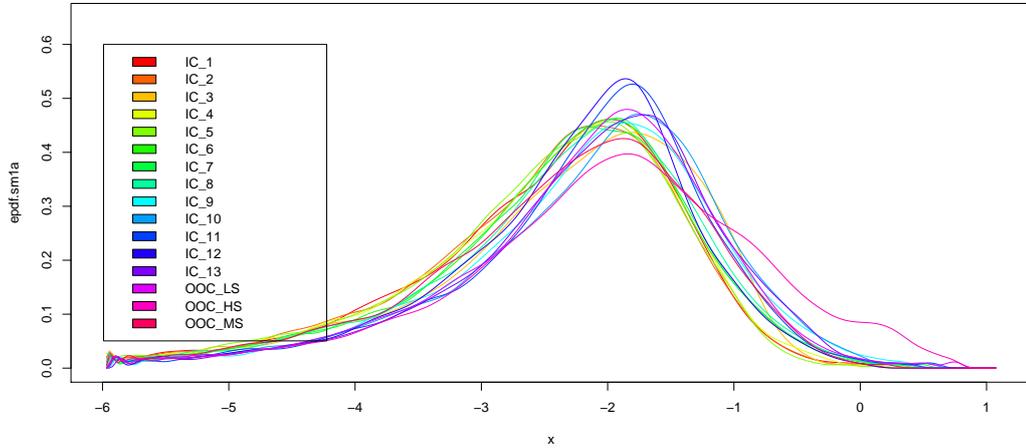


Fig.5.6: stime delle funzioni di densità f_P^i . Si noti la variabilità del campione e il comportamento unico, visibile a occhio nudo, di OOC-HS

Si osserva che anche negli elementi supposti in controllo è presente una certa variabilità, specialmente nella ampiezza e altezza del picco, che risulta localizzato in corrispondenza di distanze di circa $e^{-2} \approx 0.135mm$. Inoltre, risulta ancora più evidente il forte scostamento della funzione di densità OOC-HS rispetto al resto del campione. Possiamo fornire un esempio di applicazione delle operazioni di somma e prodotto in $B(K)$ mostrando il grafico della funzione di densità media

$$\bar{f}_P = \frac{1}{N} \odot \bigoplus_{j=1}^{16} f_P^j.$$

Ovviamente, \bar{f}_P deve essere normalizzata, pertanto il suo calcolo richiede un passaggio di integrazione numerica volto ad ottenere il valore della costante di normalizzazione

$$\int_C \left(\prod_{j=1}^{16} f_P^j \right)^{\frac{1}{N}}.$$

Mostriamo in Fig. 5.7 il grafico della densità media, rispetto alla quale tutte le funzioni di densità verranno centrate.

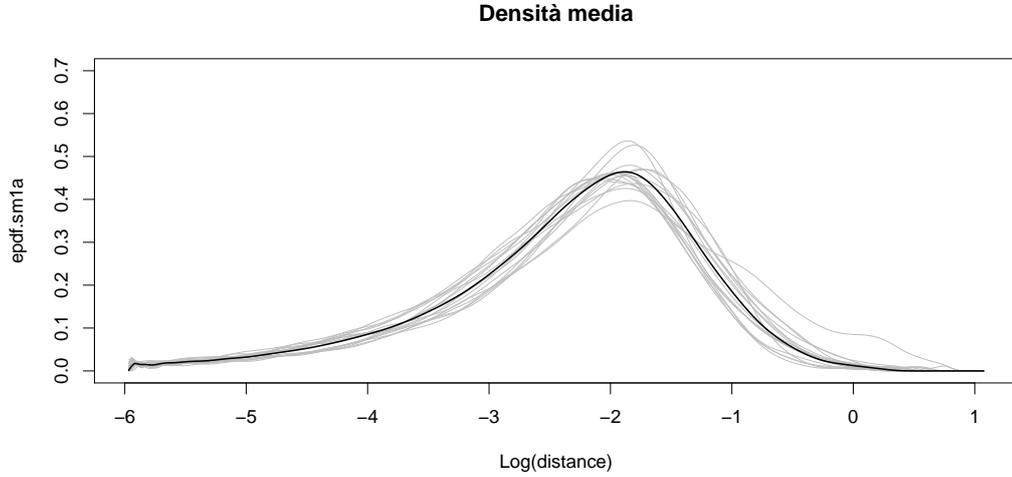


Fig.5.7: funzione di densità media \bar{f}_P (in nero) sovrapposta al campione f_P (in grigio).

5.2.2 Calcolo delle componenti principali e tipologie di controllo

Gli stimatori adottati per le f_P^j ci consentono di valutare in ogni punto dell'intervallo C tali densità. Pertanto la SFPCA può essere condotta senza specificare una base di $L_0^2(C)$, seguendo i seguenti passi:

1. Annullamento della media del campione e calcolo dei trasformati via clr : $X_P^j = clr(f_P^j \ominus \bar{f}_P)$;
2. Discretizzazione dell'intervallo $K = [a, b]$, con una partizione equispaziata di M elementi $a = c_1 < c_2 < \dots < c_M = b$;
3. Discretizzazione degli X_P^j . In particolare ogni X_P^j viene trasformato in un vettore di \mathbb{R}^M la cui k -esima componente è uguale a $X_P^j(c_k)$. A questo punto il campione di partenza è stato trasformato in un campione multivariato di $N = 16$ elementi di \mathbb{R}^M ;
4. PCA standard sul campione multivariato. La j -esima componente principale costituisce l'approssimazione discreta di $\xi_j = clr(\zeta_j)$, dove $\zeta_j \in B^2(K)$ è la j -esima componente principale del campione di partenza;
5. Uso delle approssimazioni discrete per il calcolo di $\zeta_j = clr^{-1}(\xi_j)$.

Con questo metodo, gli autovalori λ_j e le autofunzioni ξ_j sono approssimate da autovalori e autovettori della matrice di covarianza campionaria V , definita da

$$v_{rs} = \frac{1}{N-1} \sum_{j=1}^N X_P^j(c_r) X_P^j(c_s), \quad r, s = 1, \dots, M.$$

A valle del calcolo delle componenti principali, siamo in grado di passare alla fase interpretativa ed al controllo statistico di processo. Concludiamo questa breve sezione osservando che il controllo può essere condotto secondo due approcci differenti, con differenti obiettivi:

- Controllo in “Fase 1”: si suppone di non avere dei criteri per stabilire quali elementi siano o meno in controllo, usando pertanto l'intero dataset come training set per le analisi statistiche. Sulla base di tali analisi, si isolano dal dataset utilizzato gli elementi che risultano fuori controllo secondo criteri opportuni. È evidente che ci aspettiamo, se il metodo utilizzato è abbastanza robusto, che gli elementi individuati coincidano con quelli che sappiamo essere difettosi.
- Controllo in “Fase 2”: si effettua il training sugli elementi che si suppone essere a priori in controllo, e si verifica che gli elementi difettosi vengano classificati come tali dalla carta di controllo risultante.

È evidente che un metodo abbastanza robusto da superare la Fase 1, ovvero che individua i pezzi difettosi anche quando essi sono stati usati nella fase di training, risulterà adatto al controllo in Fase 2. Pertanto, ci concentreremo principalmente sul controllo in Fase 1, effettuando dunque il calcolo delle componenti principali tenendo conto di tutte e 16 le funzioni di densità a nostra disposizione. Dedicheremo tuttavia una breve sezione al controllo in Fase 2, conducendo gli opportuni confronti.

5.2.3 Analisi della varianza spiegata, proiezioni, interpretazione delle autofunzioni

In Fig. 5.8 si riporta il grafico della varianza cumulata, calcolata attraverso gli autovalori della matrice V definita nella sezione precedente. L'ispezione del grafico è necessaria per poter scegliere il numero B di componenti di cui tener conto nella costruzione delle carte di controllo.

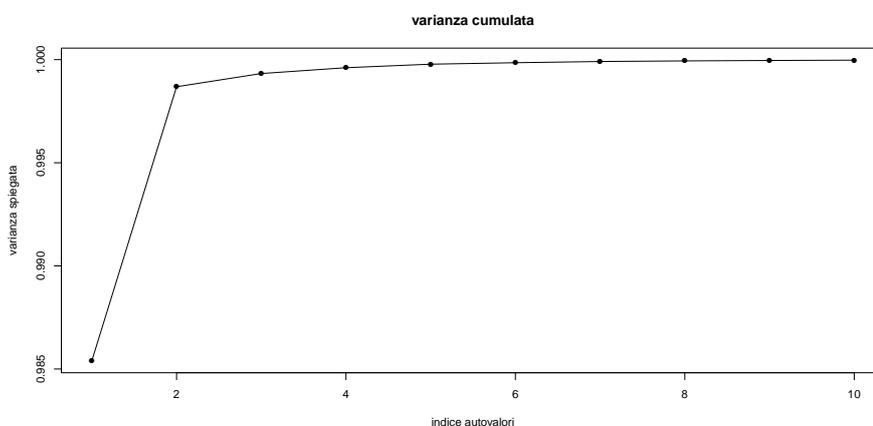


Fig.5.8: grafico della varianza cumulata, si nota il tipico "gomito" in corrispondenza del secondo autovalore

Le prime due componenti spiegano oltre il 99% della variabilità, pertanto tutte le analisi che seguono sono state svolte tenendo conto esclusivamente delle componenti ζ_1 e ζ_2 . Mostriamo in Fig. 5.9 le proiezioni del dataset sul sottospazio di B^2 generato da tali componenti (ovvero, $\forall j = 1 \dots, N$ facciamo corrispondere a f_P^j la proiezione $z_{j1} \odot \zeta_1 \oplus z_{j2} \odot \zeta_2$).

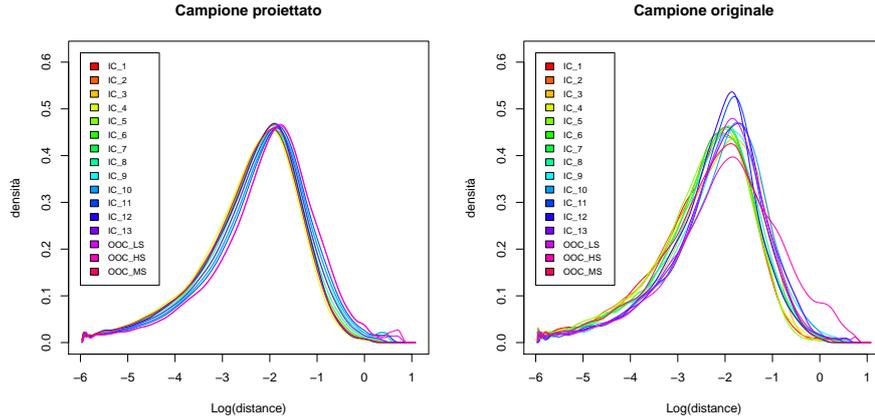


Fig.5.9: proiezioni a sinistra, dataset originale a destra, proiezione sulle prime due componenti.

Dalle proiezioni notiamo un fatto nuovo, ovvero che gli elementi difettosi OOC-LS e OOC-HS (in violetto e fucsia, rispettivamente) presentano una pronunciata tendenza alla multimodalità, con un secondo picco situato in corrispondenza distanze più elevate, quando proiettati sul sottospazio scelto. Tale tendenza risulta molto più pronunciata se ci si limita a effettuare la proiezione sul sottospazio monodimensionale generato da ζ_1 (Fig.5.10).

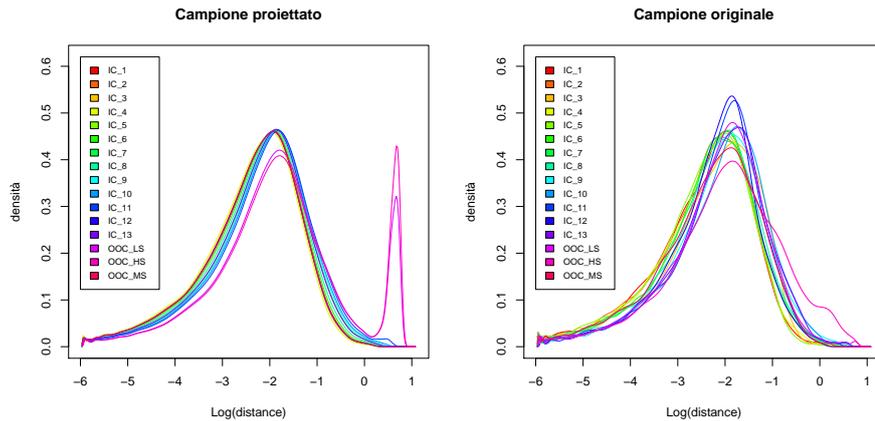


Fig.5.10: proiezioni a sinistra, dataset originale a destra, proiezione sulla prima componente.

In questo caso, è evidente un forte scostamento dei pezzi difettosi rispetto al campione. Si può apprezzare la potenza del metodo osservando che, mentre nel campione di partenza solo OOC-HS sembra distinguersi

nettamente, dopo aver effettuato la proiezione anche l'elemento OOC-LS, caratterizzato da difetti meno gravi, appare fortemente differenziato. Il profilo delle proiezioni di entrambe le densità corrispondenti ai pezzi fuori controllo risulta completamente diverso da quello degli altri elementi che compongono il dataset, confermando l'idea che le densità corrispondenti a pezzi difettosi hanno in effetti un comportamento *geometrico* molto diverso rispetto alle altre. Pertanto, anche limitandosi ad una semplice ispezione qualitativa dei risultati della SFPCA si riesce a separare in maniera piuttosto netta i pezzi difettosi dai pezzi in controllo. Per poter interpretare in maniera più precisa i risultati, ricorriamo ad altri due strumenti grafici (Ramsay e Silverman 2005): gli scatterplot degli scores z_{jk} lungo le componenti e i grafici della media perturbata lungo le direzioni delle componenti principali. Il grafico in Fig.5.12, che riporta gli scostamenti di 1.5 deviazioni standard

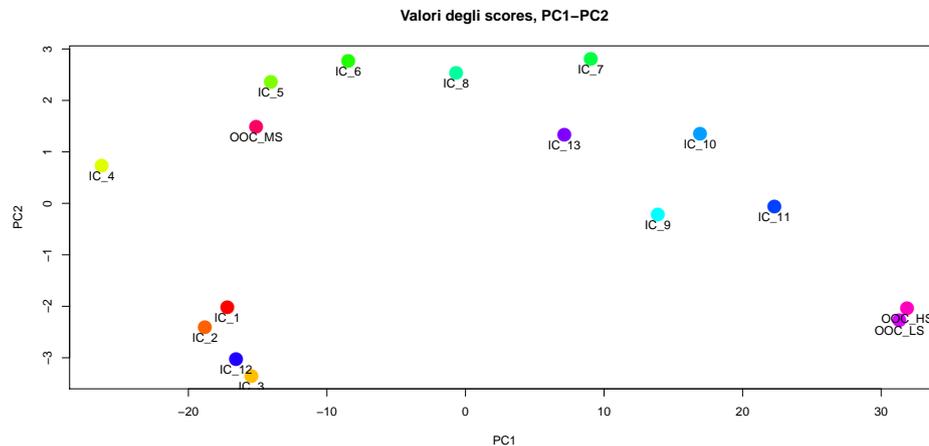


Fig.5.11: Scatterplot degli scores lungo le componenti (prima-seconda). Si noti l'evidente shift positivo dei fuori controllo lungo la direzione della prima componente

dalla media lungo le prime due componenti, ci consente di fornire una precisa interpretazione delle componenti principali, o quantomeno della prima componente. Si tratta di un fatto di notevole importanza, in quanto è ben noto che in ambito di Controllo Statistico di Processo i risultati dell'analisi delle componenti principali risultano spesso estremamente difficili, se non impossibili, da interpretare (Colosimo e Pacella 2007), fatto che costituisce il principale ostacolo all'applicazione di tale famiglia di metodi. D'altro canto, nel caso di densità basate su distanze, si conosce con precisione il significato delle variabili aleatorie di cui gli elementi del dataset sono densità (sezione **3.2**). Pertanto, da uno sguardo al sotto-grafico di sinistra possiamo subito dedurre che tutte le densità che presentano un forte shift positivo rispetto alla media, lungo la direzione di ζ_1 , hanno proiezioni che presentano un picco anomalo in corrispondenza di distanze relativamente elevate, associabili dunque a parti non conformi. Gli elementi del dataset con shift negativo sembrano discostarsi molto poco dalla media, e in ogni caso hanno tendenza

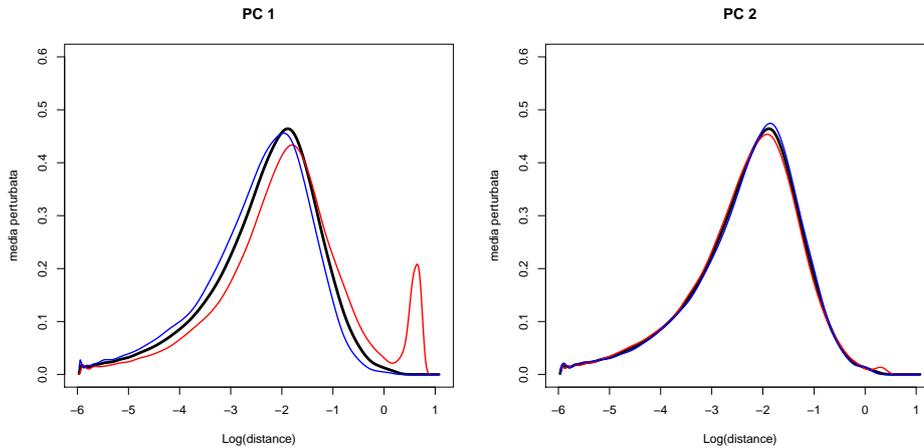


Fig.5.12: A destra densità media (nero) perturbata da $1.5\sqrt{\lambda_1}\zeta_1$ (rosso) e $-1.5\sqrt{\lambda_1}\zeta_1$ (blu). A sinistra la densità medià viene perturbata da $1.5\sqrt{\lambda_2}\zeta_2$ (rosso) e $-1.5\sqrt{\lambda_2}\zeta_2$ (blu)

a favorire distanze relativamente basse (sono molto vicine al prototipo). Dai grafici degli scatterplot in Fig.5.11 si può trovare riscontro di quanto osservato finora: nel grafico di sinistra in particolare, in cui in ascissa sono presenti gli scores relativi alla prima componente e in ordinata quelli relativi alla seconda. Si vede subito che i due elementi fuori controllo presentano un forte scostamento nel verso positivo della direzione ζ_1 . Nel proporre una carta di controllo, sarebbe dunque opportuno tenere conto dell'interpretazione dei risultati. Ci concentriamo dunque sulla carta T^2 , che per definizione rileva scostamenti anomali dalla media *lungo le componenti selezionate*, anche se non tiene conto del verso in accordo al quale tali scostamenti si verificano.

5.2.4 Carta T^2 in Fase 1 e 2, possibili adattamenti, conclusioni

Mostriamo in Fig. 5.13 la carta T^2 relativa alle analisi precedenti, che ricordiamo essere state condotte in “Fase 1”. Ricordiamo che

$$T_j^2 = \sum_{k=1}^B \frac{z_{jk}^2}{\lambda_j}, \quad j = 1 \dots N$$

e che il numero B di componenti usate è pari a 2. Si osservi che gli elementi fuori controllo OOC-HS e OOC-LS appaiono ben distinti dagli altri e che, come già osservato in precedenza, OOC-MS risulta indistinguibile dagli elementi in controllo, poiché stiamo facendo riferimento al dataset \mathbf{f}_P . Date le ridotte dimensioni del campione, ha poco senso costruire un limite di controllo con quantili empirici, ma possiamo comunque avere un'idea della variabilità della statistica T^2 attraverso un boxplot, riportato ancora in Fig 5.13.

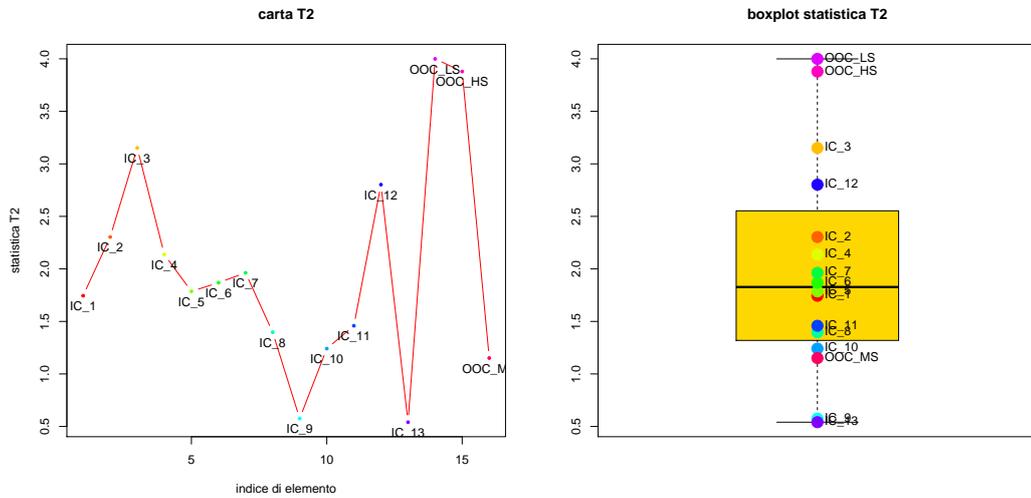


Fig.5.13: statistica T2 per ogni elemento del campione, in forma di carta di controllo e di boxplot

Il fatto che si verifichi una netta identificazione, in Fase 1 e con solamente 16 campioni disponibili è un risultato promettente. In effetti, lasciando inalterata la numerosità del campione ma effettuando l'analisi in fase 2, ovvero calcolando le componenti ζ_j usando (ad esempio) i primi 10 dei 13 componenti in controllo, si ha il risultato riportato nella Fig. 5.14.

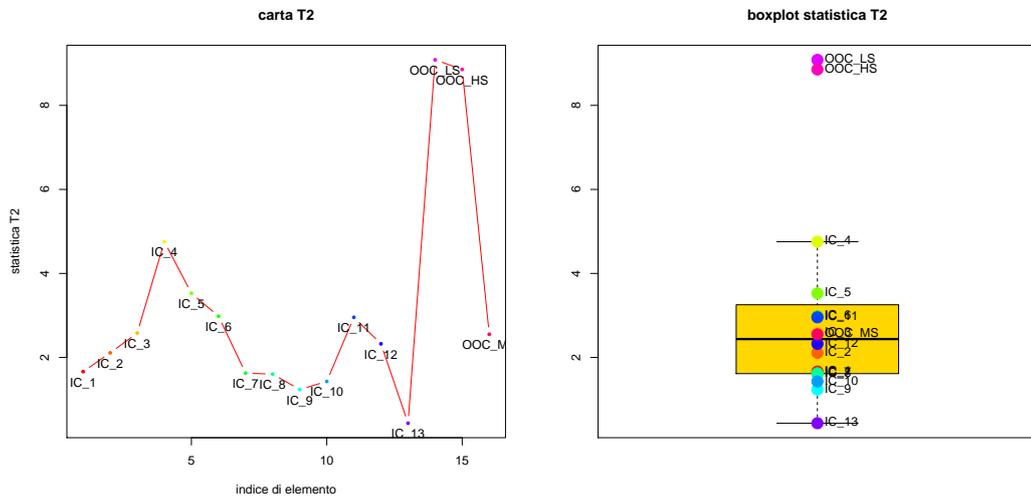


Fig.5.14: statistica T2 per ogni elemento del campione, Fase 2, in forma di carta di controllo o di boxplot

In Fase 2, pertanto, i fuori controllo presentano statistica T^2 ben lontana dall'ultimo quartile empirico, mentre gli elementi in controllo IC_i , $i = 11, 12, 13$, non utilizzati nella fase di training, non presentano comportamenti anomali. Lo stesso si può dire di OOC-MS. Concludiamo la sezione osservando che l'adozione della carta T^2 produce risultati promettenti pur

non utilizzando a pieno la conoscenza che deriva dall'interpretazione della prima componente principale fornita nella sezione precedente. In effetti, pur dando peso a qualunque scostamento di un elemento del campione lungo determinate *direzioni*, la carta T^2 non tiene in alcun conto il *verso* (positivo o negativo) di tali scostamenti. Ad esempio, abbiamo già concluso che a scostamenti in verso negativo lungo la prima componente corrispondono elementi particolarmente simili al prototipo, e comunque non molto diversi dalla media (Fig. 5.12), mentre a scostamenti in verso positivo corrispondono inevitabilmente difetti più o meno gravi. Si potrebbe immaginare di incorporare tale conoscenza nel meccanismo di controllo, proponendo una carta che tenga conto anche del significato delle autofunzioni. Nel nostro particolare caso, si potrebbe dare un peso maggiore agli spostamenti positivi lungo la prima componente rispetto a quelli negativi. In ogni caso, non rientra nello scopo della tesi produrre varianti che si adattino solo a questo particolare caso studio, e preferiamo dunque soffermarci sulle due caratteristiche che riteniamo essere i migliori pregi del metodo illustrato:

- Si tratta di un metodo *generale*, applicabile a qualunque dataset di mesh, nei limiti delle risorse computazionali a disposizione per il calcolo delle mappe di distanza.
- È un metodo che produce risultati *interpretabili*. Dato il preciso significato delle mappe di distanza, e dunque dei supporti delle funzioni di densità, è difficile immaginare una situazione in cui sia impossibile produrre un'interpretazione ragionevole almeno riguardo al significato della prima componente principale, adattando di conseguenza il design del meccanismo di controllo.

Ci soffermeremo ulteriormente sull'analisi critica del metodo, quando condurremo gli opportuni confronti con il metodo spettrale che verrà introdotto nei capitoli 6 e 7.

5.3 Analisi del dataset \mathbf{f}_S

L'analisi del dataset \mathbf{f}_S sarà del tutto analoga a quella del dataset \mathbf{f}_P appena conclusa. L'unica differenza consiste nel diverso significato geometrico del dataset: ci aspettiamo infatti la classificazione come difettoso dell'elemento OOC-MS, e soltanto di esso. Pertanto, l'estensione della sezione sarà ridotta rispetto alla precedente, e verranno mostrati solo i grafici fondamentali alla comprensione dei risultati. A meno che venga diversamente specificato, tutta l'analisi è svolta in Fase 1. Mostriamo in Fig. 5.15 il grafico della varianza spiegata relativo alla SFPCA sul nuovo dataset. Anche in questo caso, la prima componente principale spiega la maggior parte della varianza, mentre mantenendo le prime due si raggiunge un livello di varianza spiegata superiore al 99%. Usiamo pertanto le prime due componenti nell'analisi che segue.

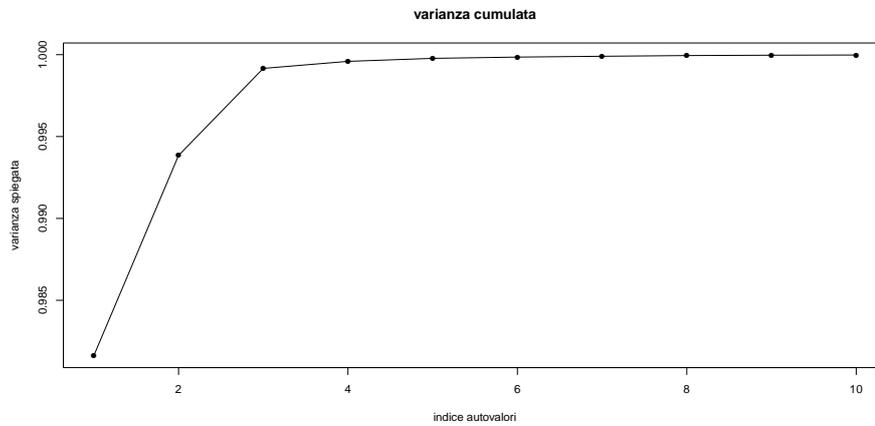


Fig.5.15: grafico della varianza cumulata

In analogia al caso precedente, visualizziamo le proiezioni del campione sul sottospazio generato da ζ_1 e ζ_2 (Fig.5.16) e sul sottospazio generato da ζ_1 (Fig.5.17). Si nota immediatamente come l'elemento difettoso OOC-MS

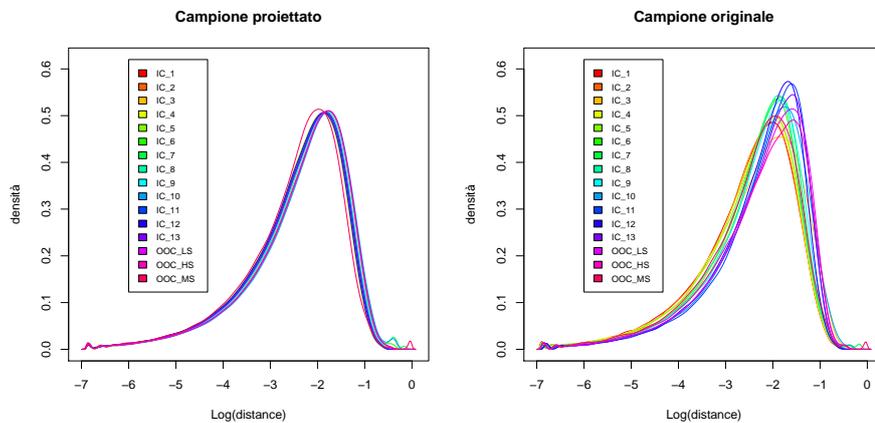


Fig.5.16: proiezioni a sinistra, dataset originale a destra, proiezione sulle prime due componenti.

si differenzi in maniera netta quando proiettato sulla prima componente principale. Ci si può rendere conto di questo fatto osservando lo scatterplot degli scores lungo le prime due componenti, in associazione ai grafici della media perturbata (Fig.5.18 e 5.19).

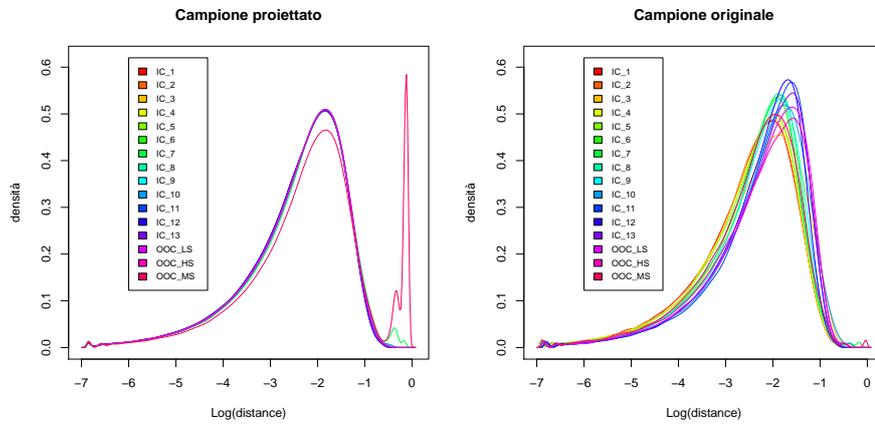


Fig.5.17: proiezioni a sinistra, dataset originale a destra, proiezione sulla prima componente.

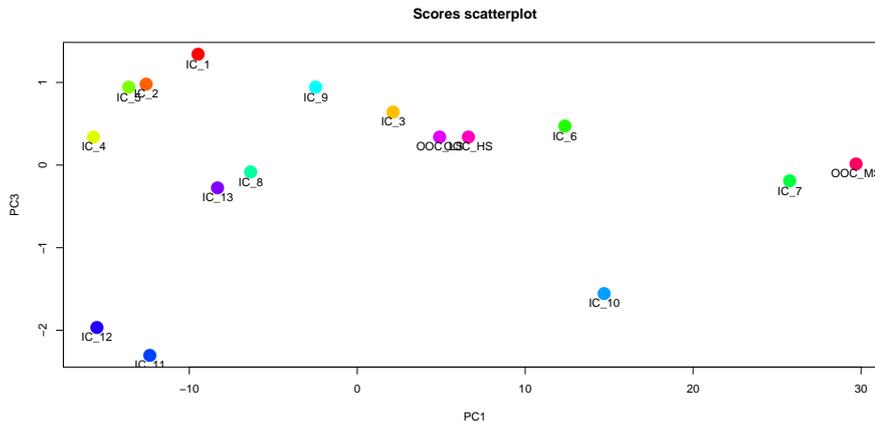


Fig.5.18: Scatterplot degli scores lungo le prime due componenti. Si noti l'evidente shift positivo di OOC-MS lungo la direzione della prima e della seconda componente

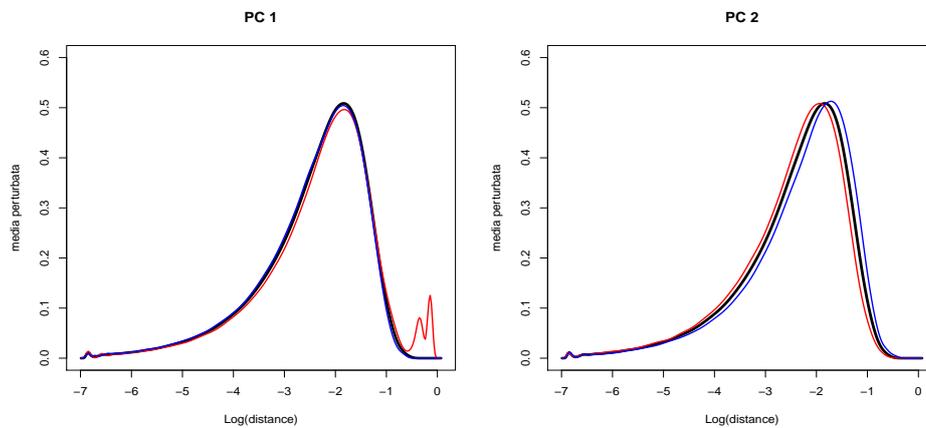


Fig.5.19: A destra densità media (nero) perturbata da $1.5\sqrt{\lambda_1}\zeta_1$ (rosso) e $-1.5\sqrt{\lambda_1}\zeta_1$ (blu). A sinistra viene rappresentata la stessa operazione, sulla seconda componente.

Possiamo dare, anche in questo caso, una differente interpretazione agli shift positivi e negativi lungo la prima componente, ed è molto facile intuire che il picco presentato dalla densità associata a OOC-MS proiettata sulla prima componente corrisponda alla massa di punti mancante (il trabecolo). In questo caso la carta T^2 fornisce indicazioni molto chiare, con una netta individuazione del pezzo difettoso, anche in Fase 1. Si direbbe che il metodo sia abbastanza sensibile anche a porzioni di mesh anomale relativamente piccole (Fig 5.22)

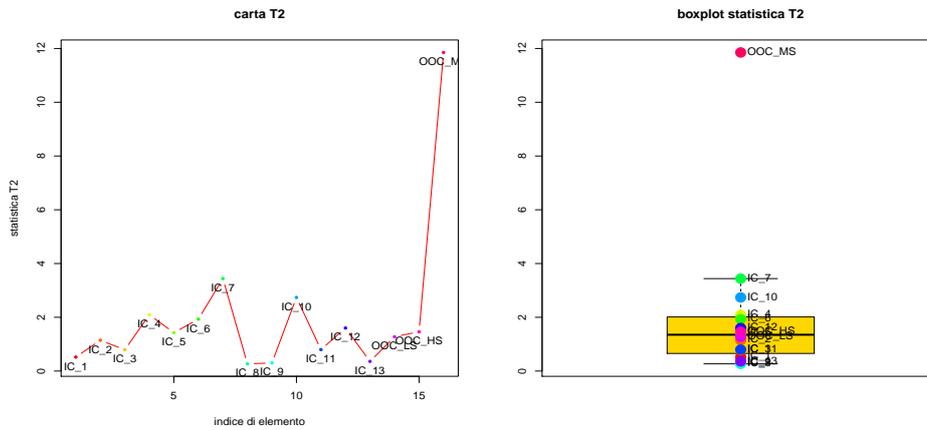


Fig.5.20: Carta T^2 e boxplot relativi alla Fase 1, si nota subito il comportamento anomalo di OOC-MS

Anche in fase 1 e con la scarsa numerosità del campione a disposizione, la statistica T^2 relativa all'elemento difettoso si posiziona ben oltre l'ultimo quartile empirico, e i risultati ottenuti effettuando il training in fase 2 sono del tutto analoghi. L'irregolarità nella struttura trabecolare rende l'elemento difettoso un "estraneo" rispetto al resto del campione, almeno dal punto di vista del comportamento geometrico della corrispondente densità di probabilità all'interno dello spazio B^2 . Viene inoltre confermato che l'analisi di f_S non rileva alcuna anomalia relativa a OOC-HS e OOC-LS, come ci si aspettava. Tutti i risultati ottenuti sono dunque coerenti con l'intuitivo significato geometrico alla base della definizione delle mappe di distanza introdotte nel capitolo 3.

Capitolo 6

Operatore di Laplace-Beltrami e analisi di superfici

6.1 Introduzione e cenni di geometria riemanniana

Ci proponiamo adesso di affrontare il problema dell'analisi di un campione di superfici da una prospettiva differente. Gli strumenti illustrati nei precedenti capitoli sono stati sviluppati unicamente in funzione di una possibile applicazione ad un contesto produttivo, e pertanto hanno come punto di partenza una rappresentazione del dataset che è il più possibile priva di informazioni non necessarie alla soluzione dello specifico problema del Controllo Statistico di Processo relativo a difetti di natura geometrica (assenza di materiale, abbondanza di materiale, distorsione geometrica di parte della mesh). Tale approccio, come osservato nel capitolo precedente, presenta notevoli punti di forza, quali una relativamente semplice interpretazione dei risultati ed un'ottima generalità, derivante da una grande flessibilità rispetto alle caratteristiche delle mesh che compongono il dataset di partenza. L'interpretabilità del metodo, in particolare, è una caratteristica preziosa in ogni contesto. È anche presente, tuttavia, qualche limitazione che merita di essere menzionata.

- Ogni elemento S_i del dataset viene rappresentato da una coppia di densità, ma tale rappresentazione è dipendente dal prototipo P . Se si immagina che, per un qualche motivo, il prototipo subisca delle modifiche, o se ci si dovesse accorgere che la mesh relativa al prototipo non lo riproduce in maniera sufficientemente accurata, le rappresentazioni di ogni componente del dataset diventerebbero inutilizzabili, e l'intera fase di allineamento delle mesh e training dovrebbe essere ripetuta. In termini più teorici, la rappresentazione *non è intrinseca*,

in quanto non dipende solamente dalle caratteristiche dell'elemento preso in considerazione, bensì anche dalle caratteristiche del prototipo (e anche dall'accuratezza ottenuta nella fase di allineamento).

- La riduzione dei supporti delle densità ad un supporto unico è un passaggio critico, la cui ragionevolezza deve essere valutata caso per caso, in quanto essenzialmente basata sull'ipotesi che i difetti siano “piccoli” rispetto all'effettiva estensione delle mesh.
- Come già osservato, è imprescindibile una fase di allineamento la cui automatizzazione è decisamente non banale. Questo fatto è intimamente legato al primo punto: il fatto che la rappresentazione non sia intrinseca la rende dipendente dalla collocazione che l'elemento rappresentato ha nello spazio euclideo. Volendo usare una terminologia più precisa, la rappresentazione di un elemento X , oltre a dipendere, com'è ragionevole, dall'immersione $i : X \rightarrow \mathbb{R}^3$, che ne codifica tutte le caratteristiche, cambia ogni qual volta che tale immersione viene composta con un'isometria (una rototraslazione). É dunque evidente che lo sviluppo di una rappresentazione alternativa, invariante per isometrie, renderebbe il passaggio di allineamento del tutto superfluo.

Vedremo a breve come sia possibile ottenere una rappresentazione intrinseca degli elementi del dataset, attraverso la decomposizione spettrale del relativo operatore di Laplace-Beltrami (Reuter et al. 2006). Tale rappresentazione non deriverà dal contesto applicativo in cui pensiamo di utilizzarla, bensì da risultati matematici che rendono lo spettro del Laplaciano una fonte di informazioni estremamente ricca. Non è un caso, del resto, che le tecniche spettrali siano ben consolidate nell'ambito della Computer Vision, e che si siano dimostrate, in tempi molto recenti, anche adatte al Controllo Statistico di Processo, come già anticipato nell'Introduzione. Prima di addentrarci nel campo della geometria riemanniana, necessaria per comprendere come la definizione dell'operatore di Laplace possa essere estesa dal semplice caso di \mathbb{R}^n a quello delle superfici, è bene soffermarsi su alcune considerazioni di carattere intuitivo, che verranno approfondite ulteriormente nella sezione **6.2**. L'operatore di Laplace, nella sua forma più nota, associa ad una funzione reale di \mathbb{R}^n due volte differenziabile la somma delle sue derivate parziali seconde, o analogamente, la divergenza del gradiente della funzione stessa:

$$\Delta f = \sum_{i=1}^n f_{x_i x_i} = \operatorname{div}(\operatorname{grad}(f))$$

É ben noto che tale operatore è un elemento fondamentale di molte delle più studiate equazioni della fisica matematica. In effetti, esso può essere descritto come l'operatore che governa i fenomeni di diffusione e vibrazione su un dominio (si pensi alle equazioni del calore, delle onde e di Laplace). Come

approfondiremo in sezione **6.2**, le soluzioni di tale equazioni sono intimamente legate alla *decomposizione spettrale* dell'operatore, ovvero ai suoi autovalori e autofunzioni. É pertanto naturale supporre che gli autovalori dell'operatore, determinando una grande varietà di fenomeni fisici sul dominio, siano anche legate alle proprietà geometriche del dominio stesso. Approfondendo il problema con gli strumenti della geometria differenziale, si scopre in effetti che tale legame è così profondo che è ragionevole trattare l'insieme di autovalori λ dell'operatore su una superficie come una “firma” della superficie stessa, di modo che lo studio della variabilità di un dataset di successioni di autovalori risulta equivalente a studiare la variabilità del dataset di superfici da cui le successioni provengono, considerazione alla base di tutti i metodi statistici basati sull'analisi spettrale dell'operatore di Laplace.

Per poter definire in maniera formale l'operatore per superfici generiche, si rende necessario introdurre numerosi strumenti di geometria differenziale e riemanniana, motivo per cui il resto della sezione **6.1** è particolarmente denso di terminologia e risultati provenienti da questa branca della matematica. Non è in ogni caso necessario conoscere in profondità o comprendere ogni dettaglio matematico della sezione: basandosi sulle considerazioni intuitive appena condotte, e sui risultati e osservazioni presenti in sezione **6.2**, è possibile convincersi senza sforzo delle motivazioni che rendono il Laplaciano così importante nello studio delle superfici.

6.1.1 Prerequisiti

Sarebbe superfluo, oltre che prolisso, iniziare dai concetti basilari che costituiscono il fondamento della Geometria Differenziale, anche perchè tali concetti vengono ampiamente trattati in un gran numero di testi (e.g, Lee 1997, 2018; Godinho e Natário 2014). Preferiamo dunque concentrarci sulla presentazione dei principali risultati e definizioni tipici della geometria riemanniana, che conducono alla definizione ed all'analisi delle proprietà dell'operatore di Laplace-Beltrami. Ovviamente, anche tali risultati possono essere studiati su uno dei molti testi a disposizione, ma la loro presentazione e comprensione è certamente propedeutica alla comprensione delle motivazioni profonde che rendono l'operatore di Laplace un “privilegiato” tra gli operatori differenziali, per cui preferiamo dedicarvi uno spazio più ampio. Le seguenti nozioni di base, che possono essere approfondite in uno dei testi citati, verranno date per note:

1. Le definizioni di varietà topologica, carta locale, atlante differenziabile, varietà differenziabile, dimensione di una varietà, coordinate locali, varietà con bordo;
2. Il concetto di differenziabilità per funzioni tra varietà differenziabili, nonché la definizione di derivazione per funzioni $C^\infty(M)$, ovvero per funzioni lisce, a valori reali, aventi per dominio una varietà M ;

3. Le definizioni e le proprietà degli spazi tangente, cotangente, e del fibrato tangente;
4. I concetti di differenziale di una funzione liscia tra varietà, di campo vettoriale, di derivazione covariante di un campo vettoriale e di connessione affine;
5. Per finire, è opportuno avere familiarità con il concetto di campo tensoriale (almeno del secondo ordine), di pull-back di un campo, di base di un campo tensoriale, e con i principali risultati relativi all'algebra esterna per varietà differenziabili e alle K -forme (campi tensoriali K -covarianti e antisimmetrici), e ai loro differenziali.

Facciamo notare al lettore che tutti i concetti elencati sono indipendenti dal fatto che sulla varietà differenziabile (la superficie) su cui si lavora sia definita o meno una metrica. Si tratta infatti di concetti di calcolo differenziale astratto, e tra di essi non figurano le ben note definizioni degli operatori differenziali comuni (gradiente, divergenza, laplaciano...). Tali definizioni, fondamentali alla modellazione di qualunque fenomeno fisico si desidera studiare sulla superficie, non possono essere date a meno che non venga definita una *metrica riemanniana*, concetto che presenteremo nella prossima sezione.

6.1.2 Fondamenti di geometria riemanniana, connessione di Levi-Civita

Il lettore può fare riferimento, per ulteriori approfondimenti, al testo di (Godinho e Natário 2014), che verrà seguito per tutta la presente sezione. Definiamo il concetto di metrica riemanniana per una varietà differenziabile.

Definizione 6.1.1. *Sia M una varietà differenziabile n -dimensionale. Una metrica riemanniana su M è un campo tensoriale 2-covariante su M*

$$g : M \rightarrow T_0^2(M),$$

in cui $T_0^2(M)$ è lo spazio dei tensori 2-covarianti di M , che soddisfa le seguenti proprietà:

1. g è una funzione C^∞ tra le varietà differenziabili M e $T_0^2(M)$.
2. Per ogni $p \in M$, $g_p \equiv g(p)$ è un tensore 2-covariante in p che è anche un prodotto scalare in $T_p(M)$, lo spazio tangente a M in p

La coppia (M, g) prende il nome di varietà riemanniana.

In altre parole, una metrica riemanniana assegna per ogni p una metrica indotta da un prodotto scalare sullo spazio tangente $T_p(M)$, variando con p in

maniera liscia. Poichè ogni g_p è un prodotto scalare, è un tensore simmetrico e definito positivo, ed è pertanto rappresentabile usando le carte locali e le usuali basi dei campi tensoriali simmetrici. Se supponiamo di lavorare in un intorno coordinato, in cui sia definita una carta locale con coordinate locali \mathbf{x} , definiamo

$$g_{ij}(p) \equiv g_p(\partial_i, \partial_j) \forall p, \forall i, j = 1, \dots, n$$

in cui i ∂_i formano la base canonica n -dimensionale dello spazio tangente in p , indotta dalle coordinate locali. Dalla simmetria segue che le funzioni reali di M g_{ij} e g_{ji} sono equivalenti per ogni coppia di indici, oltre ad essere lisce per definizione di metrica riemanniana. Possiamo adesso rappresentare g usando la base locale dei tensori simmetrici (gli indici ripetuti si intendono sommati secondo la convenzione di Einstein):

$$g = g_{ii}d(x^i)^2 + 2 \sum_{1 \leq i < j \leq N} g_{ij}dx^i dx^j$$

in cui $d(x^i)^2$ è il campo tensoriale tale che $d(x^i)_p^2(v, w) = v^i w^i \forall v, w \in T_p(M)$ mentre $dx^i dx^j(v, w) = \frac{1}{2}(v^i w^j + v^j w^i) \forall v, w \in T_p(M)$, in cui le coordinate dei vettori dello spazio tangente v e w si intendono chiaramente rispetto alla base canonica dei ∂_j . La funzione a valori matriciali con elementi g_{ij} prende il nome di tensore metrico, e avrà ovviamente valori nel cono delle matrici simmetriche e definite positive. Chiameremo G il suo determinante, e g^{ij} gli elementi del suo inverso. L'aver definito una metrica permette di chiarire cosa si intende per isometria di varietà Riemanniane.

Definizione 6.1.2. *Siano (M, g) e (N, h) due varietà riemanniane. Sia $\phi : M \rightarrow N$ liscia. Diciamo che ϕ è un'isometria se:*

1. ϕ è un diffeomorfismo, ovvero è liscia, invertibile, con inversa liscia.
2. $d\phi_p : T_p(M) \rightarrow T_{\phi(p)}(N)$ soddisfa per ogni $p \in M$ la condizione di isometria

$$g_p(v, w) = h_{\phi(p)}(d\phi_p(v), d\phi_p(w)) \forall v, w \in T_p(M)$$

La capitale importanza della presenza di una metrica sulla varietà è in primo luogo attestata dal famoso Teorema di Levi-Civita, che afferma che una metrica riemanniana identifica in maniera univoca il modo in cui le operazioni di derivazione covariante dei campi vettoriali di M devono essere svolte, ovvero identifica in modo unico una connessione su M . Si osservi che questo passaggio, che può sembrare di importanza marginale, è invece assolutamente centrale: individuata la legge appropriata per svolgere le operazioni di derivazione sulla varietà, diventa automaticamente possibile modellare fenomeni fisici, ed è infatti possibile parlare di equazione del calore, delle onde, e così via. Sofferamoci dunque sul Teorema di Levi-Civita.

Teorema 6.1.1. *Sia (M, g) una varietà riemanniana. Allora esiste un'unica connessione ∇ su M che sia simmetrica e compatibile con g , ovvero che rispetti la regola di commutazione delle derivate seconde e la regola di Leibniz rispetto al prodotto scalare g :*

$$X(g(Y, Z)) = g(\nabla_X(Y), Z) + g(Y, \nabla_X(Z)) \quad \forall X, Y, Z \text{ campi vettoriali di } M$$

Si osservi che la condizione di compatibilità con la metrica non è nient'altro che una riscrittura estremamente generale della ben nota regola di Leibniz per il prodotto scalare di funzioni a valori in \mathbb{R}^n , ovvero

$$\frac{d}{dt}\langle \mathbf{x}(t), \mathbf{y}(t) \rangle = \left\langle \frac{d}{dt}\mathbf{x}(t), \mathbf{y}(t) \right\rangle + \left\langle \mathbf{x}(t), \frac{d}{dt}\mathbf{y}(t) \right\rangle$$

Con i coefficienti g^{ij} è anche possibile esprimere la connessione in coordinate locali, ma non affronteremo questo argomento nei dettagli, si rimanda a (Godinho e Natário 2014) per una trattazione completa. Vale invece la pena di soffermarsi sul concetto di prima forma fondamentale. Sia (M, g) una varietà Riemanniana e S una sua sottovarietà, con immersione $i : S \rightarrow M$. Definiamo su S il campo tensoriale risultante dal pull-back di g a S , ovvero il campo

$$i^*(g) : S \rightarrow T_0^2(S)$$

tale che

$$i^*(g)_s(v, w) = g_{i(s)}(di_s(v), di_s(w)) \quad \forall s \in S, \forall v, w \in T_s(S),$$

in cui d è il differenziale. Si può dimostrare che tale campo tensoriale è una metrica riemanniana per S , che si dice *indotta* da g . Tale metrica prende il nome di prima forma fondamentale. Per fissare le idee, se $M = \mathbb{R}^3$ e S è una superficie liscia con parametrizzazione $\mathbf{r}(u, v)$, la prima forma fondamentale, ovvero la metrica indotta su S dal prodotto scalare euclideo, è esattamente quella usuale, ovvero la metrica rappresentata dal tensore

$$\begin{pmatrix} \|\mathbf{r}_u\|^2 & \langle \mathbf{r}_u, \mathbf{r}_v \rangle \\ \langle \mathbf{r}_u, \mathbf{r}_v \rangle & \|\mathbf{r}_v\|^2 \end{pmatrix}.$$

Pertanto, per superfici nello spazio siamo sempre in grado di definire una struttura di varietà riemanniana. Si osservi che il determinante G del tensore metrico è uguale a $\|\mathbf{r}_u \times \mathbf{r}_v\|^2$, ed è dunque legato al valore dell'elemento d'area infinitesimo dalla formula

$$d\sigma = \sqrt{G} du dv \quad (6.1)$$

che costituisce il punto di partenza per la generalizzazione del calcolo integrale alle varietà riemanniane generiche.

6.1.3 Calcolo integrale e operatori differenziali su varietà riemanniane

Anche per questa sezione, il riferimento principale è (ibid.). La presenza di una metrica consente, oltre al calcolo delle derivate sulla varietà, anche di definire in maniera univoca la generalizzazione degli operatori differenziali e del calcolo integrale di \mathbb{R}^n . In primo luogo, è possibile fornire una semplice definizione di gradiente.

Definizione 6.1.3. *Sia (M, g) una varietà riemanniana n -dimensionale. Chiamiamo gradiente l'operatore grad che associa ad ogni $f \in C^\infty(M)$ l'unico campo vettoriale $\text{grad}(f)$ che soddisfa*

$$df_p(v) = g_p(\text{grad}(f)_p, v) \quad \forall p \in M, \quad \forall v \in T_p(M),$$

in cui in questo caso df_p è il differenziale di f espresso in forma di funzionale di $T_p(M)$.

Volendo usare la terminologia dell'analisi funzionale, il gradiente di f in p è l'elemento di Riesz del funzionale df_p rispetto al prodotto scalare g_p . Si osservi che si tratta di una generalizzazione che comprende il caso della classica definizione data in \mathbb{R}^n . La formula ci consente anche di esprimere il gradiente in coordinate. Mostriamo il procedimento, per avere un'idea di come vengono usualmente ottenute le formule in coordinate dalle corrispettive versioni astratte. Le espressioni in coordinate locali degli altri operatori verranno fornite senza dimostrazione. Abbiamo, senza esplicitare la dipendenza da p per semplicità, e con convenzione di Einstein:

$$df_p(v) = df(v^i \partial_i) = v^i \frac{\partial f}{\partial x^i}.$$

Mentre

$$g(\text{grad}(f), v) = g(\text{grad}(f)^k \partial_k, v^j \partial_j) = \text{grad}(f)^k v^j g_{kj}$$

Se v è l'elemento r -esimo della base canonica dello spazio tangente si trova subito

$$\frac{\partial f}{\partial x^r} = \text{grad}(f)^k g_{kr}$$

in cui si somma su k e r è fissato, da cui si deduce immediatamente

$$\text{grad}(f)^k = g^{kr} \frac{\partial f}{\partial x^r}$$

in cui si somma su r e k è fissato. Deduciamo che, in coordinate locali, l'espressione del campo vettoriale $\text{grad}(f)$ è

$$\text{grad}(f) = g^{kr} \frac{\partial f}{\partial x^r} \partial_k.$$

Il prossimo passo consiste nella definizione di divergenza, ma è prima necessaria una breve digressione sul calcolo integrale. Il risultato centrale è il seguente.

Proposizione 6.1.1. *Sia (M, g) una varietà Riemanniana n -dimensionale orientata. Allora esiste un'unica n -forma ω_g su M detta forma di volume, che soddisfa le proprietà seguenti:*

1. $\omega_g = \sqrt{G} dx^1 \wedge \dots \wedge dx^n$ per ogni $p \in M$ e per ogni carta locale in p , dove \wedge è il prodotto wedge tra forme
2. Per ogni p e per ogni base ortonormale positiva $\{e_1, \dots, e_n\}$ di $T_p(M)$ vale $\omega_g(e_1, \dots, e_n) = 1$

Si osservi che la formula che definisce la forma ω_g non è altro che una generalizzazione della formula (6.1) alla fine della sezione **6.1.2**. La forma di volume su una varietà permette di determinare le forme di volume su tutte le ipersuperfici immerse nella varietà stessa.

Proposizione 6.1.2. *Sia (M, g) una varietà Riemanniana n -dimensionale orientata e $S \subset M$ un'ipersuperficie regolare con immersione $\phi : S \rightarrow M$. Sia N un campo di versori normali a S . Definiamo la seguente $n - 1$ forma su M , detta prodotto interno di N e ω_g :*

$$i_N \omega_g(v_2, \dots, v_n) = \omega_g(N, v_2, \dots, v_n).$$

Allora la forma di volume su S è uguale al pull-back a S di tale forma:

$$\omega_{g_s} = \phi^*(i_N \omega_g)$$

Possiamo adesso definire, per gradi, l'integrale di una funzione scalare. Sia f una funzione scalare di M il cui supporto è compatto e contenuto in un intorno coordinato W di M , con coordinate \mathbf{x} , tali che $\mathbf{x}(W) = U \in \mathbb{R}^n$, e si consideri la n -forma $\alpha = f dx^1 \wedge \dots \wedge dx^n$. Definiamo allora

$$\int_M \alpha := \int_U f dx^1 \dots dx^n$$

Si dimostra che tale definizione non dipende dalle coordinate, e può essere estesa a k -forme a supporto compatto contenuto in M . A questo punto si generalizza immediatamente la definizione di integrale di superficie, ponendo

$$\int_M f := \int_M f \omega_g.$$

Esiste anche una generalizzazione del Teorema Fondamentale del Calcolo, ovvero il Teorema di Stokes.

Teorema 6.1.2. *Sia (M, g) una varietà riemanniana orientabile con bordo ∂M e α una $k - 1$ -forma su M . Sia $\phi : \partial M \rightarrow M$ l'inclusione di ∂M in M . Allora*

$$\int_M d\alpha = \int_{\partial M} \phi^*(\alpha)$$

Si noti che la formula ha senso perchè $d\alpha$ è una k -forma.

Vorremmo adesso dare la definizione di divergenza. Per meglio comprendere il significato di questo importante operatore, procediamo cercando di ricavare la definizione guardando al semplice caso di \mathbb{R}^n . Se $V \subset \mathbb{R}^n$ è una porzione di spazio limitata con frontiera A abbastanza regolare, e X è un campo vettoriale differenziabile di \mathbb{R}^n , il teorema della divergenza ci assicura che

$$\int_V \operatorname{div}(X) d\tau = \int_A \langle X, N \rangle d\sigma$$

con N normale esterna ad A . Pertanto, volendo generalizzare il teorema della divergenza al contesto delle varietà riemanniane, vorremmo che, per ogni campo vettoriale X di M valga la formula

$$\int_M \operatorname{div}(X)\omega_g = \int_{\partial M} g(X, N)\omega_{g\partial M}$$

in cui $g_{\partial M}$ è la metrica indotta da g su ∂M . Si può dimostrare che

$$g(X, N)\omega_{g\partial M} = \phi^*(i_X\omega_g),$$

dove

$$i_X\omega_g(v_2, \dots, v_n) = \omega_g(X, v_2, \dots, v_n)$$

per cui se applichiamo il teorema di Stokes al membro di destra, la definizione di divergenza è immediata.

Definizione 6.1.4. *Sia (M, g) una varietà riemanniana, X un campo vettoriale su M . Chiamiamo divergenza di X la funzione reale di M tale che*

$$\operatorname{div}(X)\omega_g = d(i_X\omega_g)$$

In questo modo, abbiamo ottenuto un definizione generale di divergenza, che ne preserva intatto il significato fisico. L'espressione in coordinate locali, ottenibile con un po' di calcoli, è la seguente, con convenzione di Einstein:

$$\operatorname{div}(X) = \frac{1}{G} \frac{\partial}{\partial x^k} (\sqrt{G} X^k)$$

Poichè $\operatorname{grad}(f)$ è un campo vettoriale, siamo pronti a estendere la definizione di Laplaciano.

Definizione 6.1.5. *Chiamiamo operatore di Laplace-Beltrami l'operatore Δ che ad ogni $f \in C^\infty(M)$ associa la funzione*

$$\Delta f = \operatorname{div}(\operatorname{grad}(f))$$

L'espressione in coordinate locali è in questo caso

$$\Delta f = \frac{1}{\sqrt{G}} \sum_{k,l=1}^n \frac{\partial}{\partial x^k} \left(\sqrt{G} \frac{\partial}{\partial x^l} f g^{lk} \right)$$

A questo punto, ogni equazione della fisica matematica che richieda l'uso del teorema della divergenza per essere derivata (equazione del calore, equazione di Helmholtz, equazione di Laplace e così via) può essere estesa senza difficoltà al contesto generale delle varietà riemanniane. Nella prossima sezione, dopo aver chiarito i motivi fisici che rendono l'operatore di Laplace tanto legato alle proprietà topologiche della superficie, mostreremo i principali risultati che giustificano l'adozione dello spettro di tale operatore come rappresentazione di un dato riemanniano.

6.2 Proprietà dell'operatore

Il motivo per cui l'operatore di Laplace appare così spesso nelle equazioni della meccanica dei continui è il seguente (Canzani 2013). Ricordiamo che per una funzione a valori vettoriale $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ il laplaciano è definito come la funzione a valori vettoriali la cui i -esima componente è Δf^i .

Proposizione 6.2.1. *L'operatore di Laplace-Beltrami su una varietà riemanniana è invariante per isometrie. In altre parole, data un isometria*

$$\phi : N \rightarrow M$$

abbiamo

$$\Delta(f \circ \phi) = (\Delta f) \circ \phi.$$

Inoltre, sia $F : C^\infty(\mathbb{R}^n, \mathbb{R}^n) \rightarrow C^\infty(\mathbb{R})$, invariante per rototraslazioni. Allora esistono coefficienti a_1, \dots, a_n tali che

$$F = a_j \Delta^j$$

Dalla proposizione precedente, comprendiamo per quale motivo l'operatore di Laplace è presente in ogni equazione che modelli un fenomeno naturale la cui fisica è dipendente solo dalla forma del mezzo di propagazione, e non dalla posizione che il mezzo stesso occupa nello spazio. Gli esempi principali di tali equazioni sono l'equazione del calore, l'equazione delle onde, l'equazione di Schrodinger in assenza di forze esterne. Considerazioni analoghe valgono per i fenomeni di diffusione stocastica: è un fatto ben noto che il generatore

infinitesimale del moto Browniano è proprio l'operatore di Laplace. La risoluzione classica di tutte queste equazioni è basata sulla decomposizione spettrale dell'operatore, ovvero sull'individuazione delle successioni di autovalori λ_k e di autofunzioni ψ_k che soddisfino

$$\Delta\psi_k = \lambda_k\psi_k$$

con le opportune condizioni al bordo, se la varietà su cui si lavora ne ha uno. Ottenuta la decomposizione, è possibile usare il principio di sovrapposizione per scrivere una soluzione generale alle equazioni menzionate. In altre parole, sotto opportune ipotesi di regolarità, sembrerebbe che i fenomeni di vibrazione e diffusione su una varietà siano determinati dalla successione degli autovalori λ_k , detta spettro. Questa considerazione spinge a chiedersi se la successione degli autovalori caratterizzi completamente la varietà, quesito sintetizzato in maniera estremamente comunicativa da L. Bers: “*Can one hear the shape of a drum?*” (Protter 1987). Ovvero, siamo in grado di determinare la forma di un tamburo solo ascoltando la musica che esso produce? Il quesito ha attratto e continua ad attrarre un imponente lavoro di ricerca, che cercheremo di riassumere brevemente. Per comprendere appieno il resto della sezione, è opportuno avere familiarità con i vari concetti di curvatura e di caratteristica di Eulero, argomenti classici nello studio delle varietà, per i quali si può di nuovo fare riferimento a (Godinho e Natário 2014). Per tutti i risultati enunciati nel resto della sezione, che riguardano le principali proprietà dello spettro, si è fatto riferimento a (Reuter et al. 2006).

Proposizione 6.2.2. *Sia (M, g) una varietà riemanniana compatta, e sia $\{\lambda_k\}_{k \geq 1}$ lo spettro dell'operatore di Laplace-Beltrami, con condizioni di Neumann o di Dirichlet al bordo (ovviamente, se M è senza bordo, questa differenza è irrilevante). Valgono i risultati seguenti*

1. *La successione $\{\lambda_k\}_{k \geq 1}$ è nonnegativa. Inoltre, è sempre possibile ordinare tale successione in modo da ottenere una successione non decrescente e divergente, che continuiamo a chiamare $\{\lambda_k\}_{k \geq 1}$. λ_1 è semplice e, se la varietà è senza bordo, è nullo. Con condizioni di Dirichlet al bordo, λ_1 è strettamente positivo*
2. *Se due varietà sono isometriche, allora i rispettivi operatori di Laplace hanno lo stesso spettro*
3. *Lo spettro dipende con continuità dalla metrica riemanniana definita sulla varietà (Courant e Hilbert 1963)*
4. *Data una qualunque successione finita e crescente di n numeri reali, in cui il primo elemento è nullo, esiste sempre una varietà riemanniana che ammetta la successione come parte iniziale del proprio spettro. (Verdière 1987)*

5. Come conseguenza del punto precedente, se (M, g) è incognita e S è una sottosuccessione finita dello spettro, non è possibile dedurre da S il valore di alcun autovalore che non appartenga a S , ed in particolare nessuna sottosuccessione dello spettro di una varietà ignota ne determina l'intero spettro.

Precisiamo che l'ordinamento degli autovalori può essere ottenuto seguendo un procedimento di ottimizzazione di un quoziente di Rayleigh in spazi di funzioni opportuni, come nel caso della classica Analisi delle Componenti Principali. Si veda ad esempio (Salsa 2010). D'ora in poi, tutte le successioni di autovalori si supporranno ordinate in maniera non decrescente. Se immaginiamo le varietà immerse nello spazio euclideo, osserviamo che l'invarianza per isometrie è una proprietà molto più forte della più intuitiva invarianza per rototraslazioni. Si pensi di modellare una persona come una varietà riemanniana, e di immaginare la stessa persona in piedi (M) o seduta (M'). È chiaro che non esiste una rototraslazione di \mathbb{R}^3 che trasforma M in M' . Tuttavia esiste un'isometria da M in M' . Ciò è dovuto al fatto che le isometrie agiscono *intrinsecamente*, preservando le “distanze geodetiche” tra punti corrispondenti delle varietà, mentre le rototraslazioni preservano anche le distanze euclidee. In altre parole, tutte le rototraslazioni sono isometrie, ma esistono molte isometrie che non sono rototraslazioni. Di conseguenza, una persona in piedi, seduta, distesa, mantiene invariato il proprio spettro, e questo è il motivo principale che rende i metodi spettrali così attraenti nella disciplina dell'analisi di immagini.

Passiamo adesso ad analizzare quali sono le caratteristiche di una varietà che vengono determinate dallo spettro dell'operatore. Diciamo che due varietà riemanniane sono *isospettrali* se presentano lo stesso spettro. Dal punto 2 della proposizione precedente, deduciamo che se due varietà sono isometriche, allora sono isospettrali. Il problema cruciale è, a questo punto, quello inverso, del tutto equivalente al quesito di Bers riguardante la forma di un tamburo: esistono varietà isospettrali che non siano isometriche? Prima di rispondere a questa domanda, enunciamo un teorema che ci permette di comprendere quali caratteristiche della varietà siano completamente caratterizzate dallo spettro.

Teorema 6.2.1. *Sia M una varietà riemanniana compatta. Allora lo spettro $\{\lambda_k\}_{k \geq 1}$ determina in maniera univoca le seguenti informazioni geometriche:*

1. *La dimensione di M*
2. *Il volume di M*
3. *L'integrale di curvatura scalare K di M*
4. *Se M è bidimensionale, lo spettro ne determina la caratteristica di Eulero*

5. Se M ha bordo ∂M compatto, lo spettro determina anche il volume del bordo.

É possibile dare risultati numerici molto più precisi, grazie ai numerosi risultati asintotici ottenuti nel tempo (Minakshisundaram 1953; Kac 1966; McKean e M. Singer 1967). Tali risultati vengono riassunti dal seguente teorema (Reuter et al. 2006, teorema 3).

Teorema 6.2.2. *Sia (M, g) una varietà riemanniana n -dimensionale, con spettro $\{\lambda_k\}_{k \geq 1}$. Definiamo la funzione*

$$Z(t) = \sum_{n=1}^{\infty} e^{-\lambda_n t},$$

nota in letteratura come “heat trace”. Allora Z ammette la seguente espansione asintotica per $t \rightarrow 0^+$:

$$Z(t) = (4\pi t)^{-\frac{n}{2}} \left(\sum_{i=1}^n c_i t^{\frac{i}{2}} + o(t^{\frac{n+1}{2}}) \right),$$

in cui c_0 è uguale al volume di M . Se inoltre M è compatta, con o senza bordo $d-1$ dimensionale ∂M , valgono inoltre le relazioni

$$c_1 = -\frac{\sqrt{\pi}}{2} \text{vol}(\partial M)$$

$$c_2 = \frac{1}{3} \int_M K - \frac{1}{6} \int_{\partial M} J$$

in cui K è la curvatura scalare di M e J la curvatura media di ∂M . Se in particolare M è bidimensionale e senza bordo (ad esempio una superficie chiusa),

$$c_2 = \frac{2\pi}{3} \chi,$$

in cui χ è la caratteristica di Eulero di M .

Dovrebbe essere chiaro, adesso, che lo spettro dell’operatore di Laplace contiene una notevolissima quantità di informazioni di natura topologica e geometrica. Tuttavia, la risposta alla domanda “*Can one hear the shape of a drum?*” è sfortunatamente negativa: in effetti, esistono coppie di varietà Riemanniane isospettrali ma non isometriche. Si tratta di un risultato ottenuto in tempi relativamente recenti, nel 1992 (Gordon et al. 1992). Tuttavia, l’isospettralità è un fenomeno relativamente raro, in quanto le coppie di varietà isospettrali finora costruite sono “casi patologici”, appositamente costruite per fornire controesempi. In particolare:

- Gli unici esempi di varietà isospettrali non isometriche, a due o tre dimensioni, consistono di coppie di varietà concave e con frontiera non differenziabile. Gli unici esempi di domini convessi dello spazio euclideo, che siano isospettrali ma non isometrici, sono stati formulati in dimensione almeno pari a 4.
- Non sono attualmente note triplette di varietà isospettrali.
- Nel caso di varietà bidimensionali (superfici di \mathbb{R}^3), che è esattamente il contesto in cui sono ambientati i problemi di controllo statistico che abbiamo come obiettivo, il teorema 6.2.2 assume la sua piena potenza, permettendo di dedurre praticamente tutte le informazioni topologiche relative alla varietà (volume, area, frontiera, curvatura, caratteristica di Eulero)

Per tutti questi motivi, appare del tutto inverosimile che un processo produttivo reale possa produrre coppie di oggetti rappresentabili da varietà isospettrali ma non isometriche, ed è esattamente questa osservazione che porta a considerare lo spettro, nei casi applicativi reali, come “*DNA di forma*” (ShapeDNA, Reuter et al. 2006). Un problema non banale, di grande interesse pratico, è il calcolo numerico dello spettro di una varietà, o di una mesh, in tempi ragionevoli. Si tratta di un campo che è stato, ed è tuttora, oggetto di forte interesse da parte dei ricercatori: nella prossima sezione passeremo rapidamente in rassegna alcuni metodi disponibili per la risoluzione del problema.

6.3 Soluzioni per il calcolo dello spettro

Nel lavoro di Reuters et al. (2006), oltre ai risultati teorici appena illustrati viene anche formulato un metodo per il calcolo numerico della decomposizione spettrale, basata sulla formulazione debole del problema agli autovalori con relativa condizione di Neumann o Dirichlet al bordo e successiva risoluzione approssimata attraverso il metodo degli elementi finiti, con opportuni accorgimenti nella gestione della triangolazione della superficie parametrica o della mesh in input, volti a garantire tanto una precisione di calcolo adeguata quanto tempi di esecuzione ragionevoli. Non è nostra intenzione soffermarci su questioni di calcolo numerico, ma è interessante notare come gli autori mostrino che il metodo presentato presenti buone prestazioni nell’applicazione ad immagini simulate o reali, a superfici parametriche e a mesh. Va inoltre sottolineato che, nel caso dell’applicazione a mesh, tale metodo richiede una fase di pre-processamento che può essere svolta in maniera totalmente automatica (pulizia della mesh e rimozione degli spigoli con più di due facce insistenti), eliminando dunque il problema di pre-processamento manuale emerso nella parte della tesi dedicata a SFPCA. Reuters, inoltre, ha reso liberamente fruibile l’implementazione C++ del metodo proposto nel lavoro,

sotto forma del programma ShapeDna-tria, facilmente applicabile a dataset di mesh¹. Il programma, relativamente semplice da usare, consente all'utente di scegliere l'ordine degli elementi finiti da utilizzare (dal primo al terzo ordine), il numero di autovalori da calcolare, la condizione al bordo, il livello di smoothing della mesh, ed altri parametri di minore importanza, risultando in un notevole livello di personalizzazione. Inoltre, ShapeDna-tria è già stato impiegato con successo in diversi casi di applicazione a reali dataset di immagini (si veda ad esempio Reuter 2010). Per questi motivi, nell'applicazione al dataset oggetto di studio, si è scelto tale software per il calcolo dello spettro di ogni elemento.

Tuttavia, altre proposte per il calcolo sono state avanzate nel corso degli anni. Nel lavoro (Zhao e Del Castillo 2019), che approfondiremo più dettagliatamente nel prossimo capitolo, il problema del calcolo degli autovalori viene affrontato con l'applicazione del metodo proposto in (Li et al. 2015), che evita di usare gli elementi finiti, ricorrendo piuttosto alla discretizzazione e localizzazione dell'operatore, attraverso approssimazioni successive che illustriamo brevemente. Data una superficie S nello spazio euclideo, è noto che per $t \rightarrow 0$ l'approssimazione asintotica

$$F_S^t(f)(\mathbf{x}) = \frac{1}{4\pi t^2} \int_S e^{-\frac{d_S(\mathbf{x},\mathbf{y})^2}{4t}} (f(\mathbf{y}) - f(\mathbf{x})) d\sigma(\mathbf{y}), \quad (6.2)$$

dove d_S è la distanza su S , converge puntualmente a Δf per ogni funzione reale f di S che sia due volte differenziabile. Tale approssimazione è non locale, in quanto è necessario calcolare un integrale esteso all'intera superficie per valutare l'approssimazione F_S^t in un singolo punto \mathbf{x} della superficie. Se la superficie è triangolata, ovvero si lavora con una mesh, l'approssimazione assume la forma seguente (Belkin e Niyogi 2008):

$$L_M^t(f)(\mathbf{p}_i) = \frac{1}{4\pi t^2} \sum_{T \in T_M} \frac{A(T)}{3} \sum_{\mathbf{p}_j \in V(T)} e^{-\frac{\|\mathbf{p}_i - \mathbf{p}_j\|^2}{4t}} (f(\mathbf{p}_i) - f(\mathbf{p}_j)) \quad (6.3)$$

in cui i \mathbf{p}_i sono i vertici della mesh, T_M l'insieme delle facce, $A(T)$ e $V(T)$ l'area e l'insieme dei vertici del triangolo T , rispettivamente. Da questa approssimazione è possibile ricavare immediatamente un'approssimazione finita dimensionale dell'operatore in forma matriciale, per cui si rimanda a (ibid.), la cui decomposizione spettrale costituisce un'approssimazione dello spettro del Laplaciano. In (Li et al. 2015), l'approssimazione integrale e la corrispondente versione per mesh vengono *localizzate*, considerando nel calcolo dell'integrale solo i punti della mesh all'interno di un intorno sferico S_r di \mathbf{x} :

$$F_{S_r}^t(f)(\mathbf{x}) = \frac{1}{4\pi t^2} \int_{S_r} e^{-\frac{d_S(\mathbf{x},\mathbf{y})^2}{4t}} (f(\mathbf{y}) - f(\mathbf{x})) d\sigma(\mathbf{y}).$$

¹Il lettore interessato può visitare il sito del laboratorio diretto da Reuters, <http://reuter.mit.edu/software/shapedna> per maggiori informazioni e per scaricare i software disponibili

Nello stesso lavoro, viene dimostrato che anche questa approssimazione converge puntualmente a Δf per $t \rightarrow 0$. La corrispondente espressione nel caso delle mesh è identica alla (6.3), ma estendendo la sommatoria solo a quei triangoli che si trovano all'interno dell'intorno sferico di \mathbf{p}_i . Anche in questo caso, si ottiene una versione matriciale dell'operatore (Zhao e Del Castillo 2019), di cui è possibile calcolare lo spettro. In (Li et al. 2015), viene mostrato che il metodo sembra ottenere risultati sensibilmente migliori rispetto ai metodi basati su elementi finiti quando si lavora su superfici che presentano variazioni di curvatura particolarmente brusche. Il fatto che il dataset oggetto di studio non presenti questo problema, unito alla particolare semplicità applicativa dell'implementazione, ci ha spinto a prediligere il metodo di Reuters, basato su elementi finiti. È in ogni caso evidente che, a patto di raggiungere un'accuratezza sufficiente, la scelta del metodo di calcolo è dettata unicamente dalle esigenze applicative contingenti, ed è del tutto indipendente dalla fase di analisi statistica delle successioni di autovalori.

Capitolo 7

Caso studio - Seconda Parte

7.1 Un nuovo strumento nel Controllo Statistico di Processo

Come anticipato nell'Introduzione, l'uso di tecniche spettrali basate sull'operatore di Laplace-Beltrami è piuttosto consolidato nell'ambito del riconoscimento di immagini, tanto da spingere alcuni esperti ad affermare che *“The Laplace-Beltrami operator (...) can be seen as one of the most fundamental operator in the analysis and processing of images and shapes”* (Wetzler et al. 2013, p.13), ovvero che il Laplaciano può essere considerato uno degli operatori più importanti nella disciplina del processamento di forme. Nel lavoro appena citato, il lettore potrà trovare diversi esempi che certificano quanto sia ampio il campo applicativo di tale operatore al riconoscimento di immagini. Tale vastità è senz'altro dovuta all'osservazione, già portata avanti nel capitolo precedente, che il fenomeno dell'isospettralità non isometrica è estremamente raro e sembra essere del tutto assente in reali problemi applicativi. Mantenendo un contesto generale, si può di conseguenza dire che è possibile rappresentare un dataset di immagini/mesh (X_1, \dots, X_N) , noto a meno di isometrie, facendo corrispondere ad ogni X_i la successione λ_i , che ricordiamo è ordinata in maniera non decrescente, dei primi K autovalori, ossia

$$\lambda_i = (\lambda_{i1}, \dots, \lambda_{iK})^T.$$

Si ottiene dunque il dataset $(\lambda_1, \dots, \lambda_N)$, con la ragionevole sicurezza che la perdita di informazioni sia esclusivamente dovuta al fatto che le successioni di autovalori sono troncate e quindi non possono caratterizzare totalmente le superfici (Proposizione 6.2.2). Tale proposizione consente anche di affermare che, qualora due successioni finite λ_i e λ_j presentino sensibili differenze, le superfici corrispondenti debbano altresì essere sensibilmente differenti, in quanto lo spettro dipende con continuità dalla metrica riemanniana definita sulla superficie. Tutte queste considerazioni, unite al fatto che l'uso di proprietà intrinseche delle superfici rende superfluo qualunque allinea-

mento, hanno spinto alla prima proposta di applicazione dell'operatore di Laplace-Beltrami al Controllo Statistico di Processo in campo industriale nel recentissimo lavoro di Zhao e Del Castillo (2019). Come fanno notare gli autori, le classiche applicazioni al riconoscimento di immagini consistono quasi sempre nella classificazione di un gran numero di immagini che presentano tra loro grandi differenze, un "compito che anche un uomo paziente può svolgere, in linea di principio" (tradotto da Zhao e Del Castillo 2019). Nel Controllo Statistico di Processo tuttavia, le differenze tra gli oggetti sono spesso piccole se paragonate all'estensione delle mesh, ed il nostro dataset è un esempio lampante di questo fatto. Gli autori mostrano, lavorando su modelli CAD di parti in controllo o aventi piccoli difetti, che le successioni di autovalori troncate si mostrano molto efficaci anche nell'individuazione di piccole differenze, e sono dunque utilizzabili nella progettazione di sistemi di controllo, con una buona efficacia sia in Fase 1 che in Fase 2 (carta DFEWMA, Chen et al. 2015). Le geometrie dei modelli CAD simulati utilizzati nel lavoro sono relativamente semplici, tipici esempi di output di processi di produzione classici, ed è dunque naturale chiedersi se il calcolo dello spettro si riveli ugualmente efficace nell'individuazione di pezzi difettosi, quando applicato a un dataset di oggetti con una geometria molto più complessa, provenienti da Additive Manufacturing. È a questo quesito che cercheremo di rispondere calcolando lo spettro delle mesh che compongono il nostro dataset.

7.2 Analisi del dataset oggetto di studio

Per ogni elemento del dataset, i primi $K = 100$ autovalori dell'operatore di Laplace-Beltrami sono stati calcolati, grazie al software ShapeDna-tria. Gli elementi IC6 e IC7 sono stati esclusi dall'analisi per via della presenza di irregolarità in alcune facce delle mesh, che inducono instabilità nel calcolo dello spettro (secondo i warning generati dall'esecuzione di ShapeDna-tria su tali mesh). Come già osservato, la fase di pre-processamento e pulizia delle mesh, necessaria ad una corretta esecuzione del software, può essere svolta in automatico, ad esempio usando la già menzionata libreria R `rvcg`, riducendo al minimo il tempo di manipolazione diretta delle mesh stesse. Già dall'ispezione di un semplice grafico delle successioni si ottengono interessanti risultati (Fig. 7.1). Ricordiamo che per OOC-LS e OOC-HS si intendono gli elementi con irregolarità distribuita, a lieve ed alta severità, rispettivamente. Per OOC-MS intendiamo l'elemento con un trabecolo assente.

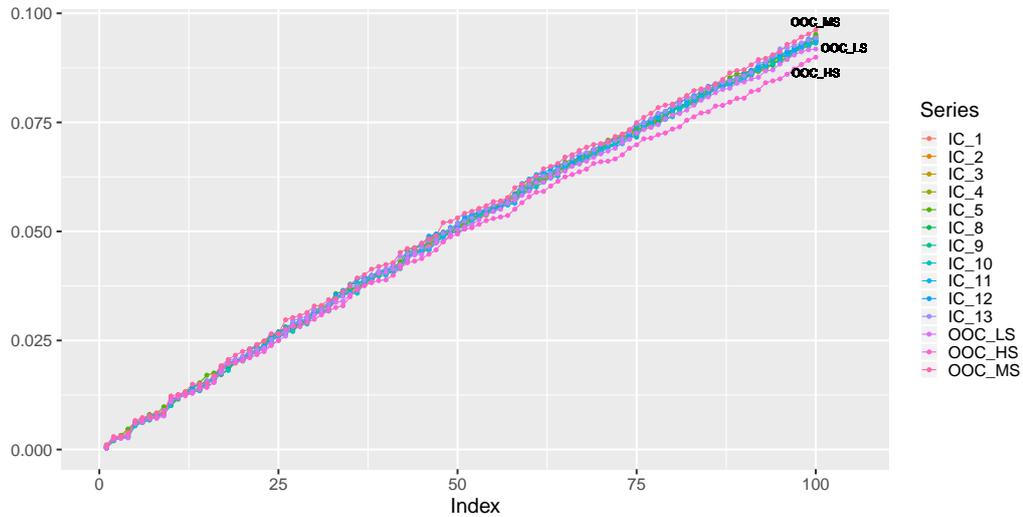


Fig. 7.1: grafico delle successioni per gli elementi del dataset, esclusi IC6 e IC7. In ordinata gli indici degli autovalori. Si noti il sistematico scostamento, visibile a occhio nudo, di OOC-HS (fucsia, in basso), OOC-MS (rosa, in alto) e OOC-LS (violetto, in basso).

La valutazione visiva dell'andamento degli spettri troncati induce naturalmente alcune osservazioni.

- Il fatto che la fase di allineamento risulti inutile è confermato. Gli spettri sono infatti calcolati su mesh non allineate, e rimangono identici se calcolati dopo che ai pezzi venga applicata qualunque rototraslazione.
- Il fatto che lo spettro dipenda con continuità della superficie (Proposizione 6.2.2) trova evidente riscontro. Tutte le successioni hanno infatti un andamento decisamente simile, ed anche gli elementi difettosi si discostano di poco (in senso assoluto) dal comportamento dei pezzi in controllo. Si tratta di un fatto ragionevole, in quanto tutti gli elementi, per quanto affetti da una certa variabilità, sono stati prodotti sulla base di uno stesso prototipo.
- Dal punto precedente deduciamo che, anche nel caso di mesh piuttosto estese (poco più di una decina di migliaia di vertici) e provenienti da oggetti con geometrie non banali, il calcolo degli autovalori con metodo degli elementi finiti risulta piuttosto stabile, in quanto produce risultati coerenti da mesh a mesh.
- La dipendenza continua trova ulteriore riscontro dal comportamento dei pezzi difettosi. I due pezzi OOC-LS e OOC-HS sono infatti affetti dallo stesso tipo di difetto, benchè con diversi livelli di severità, ed in effetti si discostano dai pezzi in controllo in maniera simile (entrambe le successioni sono spostate verso il basso rispetto alla media, con uno scostamento più evidente in corrispondenza di OOC-HS, caratterizzato

del resto da una più severa difettosità). OOC-MS, d'altro canto, si comporta diversamente da entrambi. Pertanto, è possibile apprezzare una certa coerenza tra le informazioni già a nostra disposizione e le indicazioni fornite dagli spettri.

Per avere un'idea di quanto significativi siano gli scostamenti dei pezzi difettosi rispetto alla media dei pezzi in controllo, si può fare riferimento ai boxplot autovalore per autovalore. Riportiamo la prima e l'ultima parte dello spettro, ovvero i primi trenta e gli ultimi trenta autovalori, nelle Figure 7.2 e 7.3, che per comodità sono inserite alla fine del capitolo. Con un po' di pazienza nell'osservazione dei boxplot, il lettore potrà rendersi conto che gli autovalori di OOC-MS si trovano in maniera quasi sistematica ben al di sopra del terzo quartile empirico, mentre OOC-HS si trova quasi sempre ben al di sotto del primo, con OOC-LS che si comporta in maniera molto simile, anche se meno marcata. Si tratta di un risultato incoraggiante: pur con un numero così limitato di pezzi in controllo, la semplice osservazione dei quartili empirici conferma che i pezzi difettosi assumono un comportamento peculiare distinguendosi dagli altri elementi, in maniera simile al caso della SFPCA.

La costruzione e il test dell'efficacia delle carte di controllo proposte in (Zhao e Del Castillo 2019) non risultano possibili, per via della ridotta numerosità del campione e dei pochi pezzi difettosi, in quanto la carta DEFMTWA proposta dagli autori richiede un certo numero di elementi difettosi per innescare un allarme di fuori controllo (Chen et al. 2015). Pertanto, in questo lavoro mostreremo i risultati derivanti da un approccio basato sull'analisi spettrale, su popolazioni simulate perturbando alcune delle mesh reali a nostra disposizione. In (Zhao e Del Castillo 2019) le mesh vengono perturbate aggiungendo rumore gaussiano isotropo $N(0, \sigma^2 I_3)$ alle coordinate dei vertici. Nel nostro caso, la scelta di σ^2 non è immediata, in quanto è necessario che la deviazione standard non sia tale da distruggere la struttura delle mesh, ma sia anche confrontabile con la variabilità indotta dal processo produttivo e dalla fase di scansione degli oggetti. Sarà interessante, in futuro, studiare come una mesh complessa possa essere perturbata in maniera coerente a questi elementi, nonché approfondire quale sia il rapporto tra l'intensità e il tipo della perturbazione e l'andamento della successione spettrale della mesh perturbata. Mostriamo cosa accade alle mesh con vari valori di σ , che si intende espressa in millimetri, in Fig. 7.4.

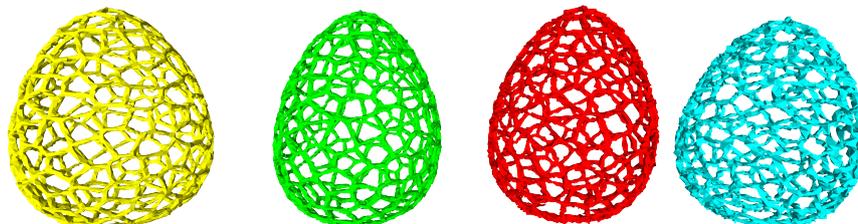


Fig.7.4: elementi perturbati a $\sigma = 0.05, 0.1, 0.2, 0.3$, da sinistra a destra

Vediamo subito che, mentre una perturbazione con $\sigma = 0.05mm$ o $\sigma = 0.1mm$ non provoca particolari alterazioni strutturali, perturbazioni più forti alterano la mesh in maniera inaccettabile. Decidiamo dunque di generare quattro popolazioni simulate, rispettivamente a partire da un elemento in controllo e dai tre elementi difettosi, applicando agli elementi di partenza rumore gaussiano con deviazione standard $\sigma = 0.1$. Per generare la popolazione in controllo, scegliamo l'elemento IC1. Riportiamo in Fig. 7.5 un esempio di popolazione simulata, mostrando i primi cinque elementi simulati in controllo.

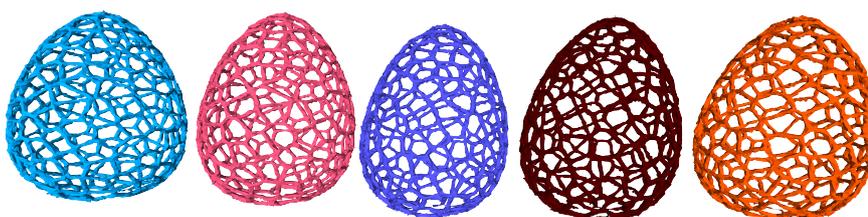


Fig.7.5: Elementi della popolazione in controllo simulata, si noti come la perturbazione non abbia alterato la struttura delle mesh

Seguendo l'esempio di (Zhao e Del Castillo 2019), generiamo $n = 10$ elementi in controllo e $m = 5$ per ogni popolazione fuori controllo, e calcoliamo lo spettro di ogni elemento così ottenuto, procedendo quindi ad un test di

permutazione volto a studiare le differenze tra i dataset così ottenuti. Prima di entrare nei dettagli del test, è interessante visualizzare gli spettri delle popolazioni simulate: in Fig. 7.6 è possibile apprezzare come le successioni appartenenti a differenti popolazioni tendano a differenziarsi.

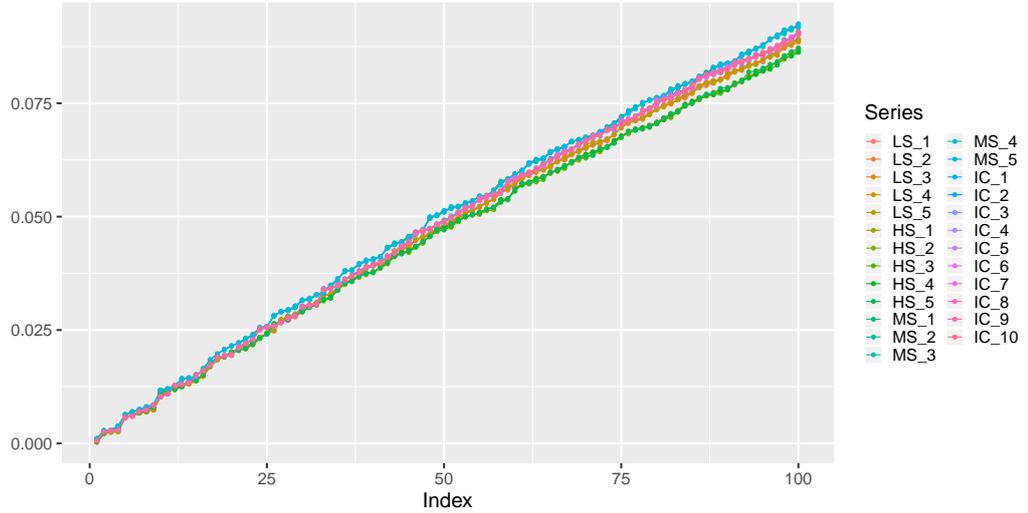


Fig.7.6: Spettri delle quattro popolazioni simulate a confronto. È immediato notare la separazione degli insiemi di successioni, analoga alla situazione degli spettri delle mesh reali. Le successioni simulate relative a OOC-MS sono in alto, dopo di che, verso il basso, abbiamo quelle relative a IC1, quindi a OOC-LS e infine quelle relative a OOC-HS.

Passiamo adesso al test di permutazione, che descriviamo facendo riferimento alla popolazione difettosa generata da OOC-HS, ma quanto illustrato si applica in maniera identica alle altre popolazioni. Sia $(\lambda_1, \dots, \lambda_n)$ il campione delle successioni spettrali corrispondenti agli elementi simulati in controllo considerati (nel nostro caso $n = 10$) e $(\lambda_{HS_1}, \dots, \lambda_{HS_m})$ l'analogo campione relativo alla popolazione generata da OOC-HS. Il test di permutazione è articolato secondo lo schema seguente:

1. Si ipotizza che il dataset multivariato $(\lambda_1, \dots, \lambda_n)$ sia costituito da elementi i.i.d. in accordo ad una legge F di media μ_0 , mentre $(\lambda_{HS_1}, \dots, \lambda_{HS_{m-1}})$ segue la legge, ma con media μ_1 . Lo scopo del test sarà valutare $H_0 : F_0 = F_1$ contro $H_1 : F_0 \neq F_1$, con una statistica test che evidenzi differenze nelle medie delle popolazioni.
2. Si calcolano le medie campionarie Λ_{IC} e Λ_{HS} del campione in controllo e di quello difettoso, rispettivamente, ovvero

$$\Lambda_{IC} = \frac{1}{n} \sum_{i=1}^n \lambda_i,$$

$$\Lambda_{HS} = \frac{1}{m} \sum_{i=1}^m \lambda_{HS_i}.$$

Si calcola quindi la statistica t_{max} (Reuter et al. 2007), definita da

$$t_{max} = \max_{1 \leq j \leq p} \frac{|\Lambda_{IC}^j - \Lambda_{HS}^j|}{SE_j}$$

in cui p è il numero di autovalori calcolati, e quindi la dimensione di ogni elemento del campione, mentre Λ_{IC}^j e Λ_{HS}^j sono la media campionaria del j -esimo autovalore nel campione in controllo e nel fuori controllo, cioè la j -esima componente di $\mathbf{\Lambda}_{IC}$ e $\mathbf{\Lambda}_{HS}$. SE_j è la stima della deviazione standard del j -esimo autovalore: siano $\sigma_{IC,j}^2$ e $\sigma_{HS,j}^2$ le varianze campionarie del j -esimo autovalore rispettivamente nel campione in controllo ed in quello difettoso, cioè

$$\sigma_{IC,j}^2 = \frac{1}{n-1} \sum_{i=1}^n (\lambda_{ij} - \Lambda_{IC}^j)^2$$

$$\sigma_{HS,j}^2 = \frac{1}{m-1} \sum_{i=1}^m (\lambda_{HS,i,j} - \Lambda_{HS}^j)^2.$$

Si definisce allora:

$$SE_j = \sqrt{\frac{(n-1)\sigma_{IC,j}^2 + (m-1)\sigma_{HS,j}^2}{m+n-2}}.$$

3. Si ripete il calcolo della statistica t_{max} permutando più volte i dataset in controllo e fuori controllo, ottenendo una stima empirica del p-value del test al punto 1 usando la frequenza con cui la statistica t_{max} calcolata sulle permutazioni risulta più elevata di quella calcolata sui dataset non permutati. Se si indica con t_{max}^* il valore della statistica sui campioni non permutati, e si compiono H iterazioni di permutazione, abbiamo la seguente stima p-value:

$$p = \frac{1}{H} \sum_{h=1}^H \mathbb{1}(t_{max}^h > t_{max}^*)$$

Mostriamo nelle Fig 7.7 e 7.8, il valore p-value del test e le distribuzioni delle statistiche t_{max} , per ognuna delle tre tipologie di difetto, valutate su 10000 permutazioni.

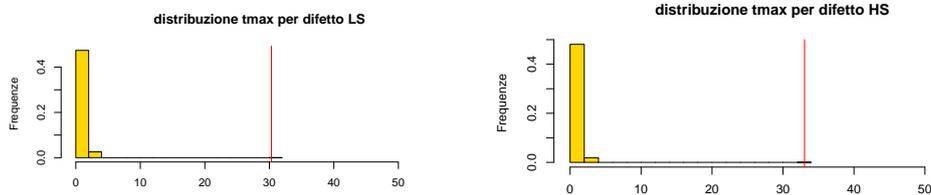


Fig. 7.7: istogrammi delle t_{max} nei casi di difetto OOC-LS e OOC-HS. In rosso il valore di t_{max} per i dataset non permutati. In entrambi i casi, il p-value del test è praticamente nullo

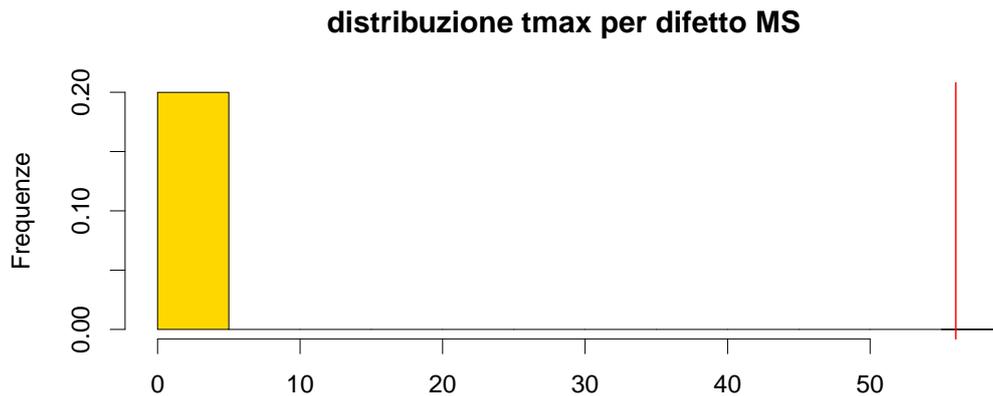


Fig.7.8: istogrammi di t_{max} per il difetto OOC-MS, con in rosso il valore nel caso dei dataset non permutati. Anche in questo caso, il p-value del test è praticamente nullo

Come si vede dai grafici, OOC-HS e OOC-LS hanno un comportamento simile, mentre la statistica t_{max} nel caso di OOC-MS è molto più alta e distante dalla distribuzione dei valori della statistica nelle permutazioni. In ogni caso, in tutti gli esempi di difetto l'ipotesi nulla è decisamente respinta, una conferma quantitativa delle impressioni che provengono dall'osservazione dei boxplot degli autovalori: sembra che gli spettri costituiscano una rappresentazione adeguata a rilevare anche piccoli difetti in dataset di oggetti complessi. Pertanto, appare concreta la possibilità, che andrà in futuro testata con studi su dataset di oggetti reali più numerosi o su simulazioni calibrate in modo da generare una variabilità confrontabile con quella di un processo produttivo, di costruire sistemi di controllo di produzione in cui è

assente qualunque fase di allineamento spaziale, anche nel caso di elementi molto complessi, eventualmente provenienti da processi AM.

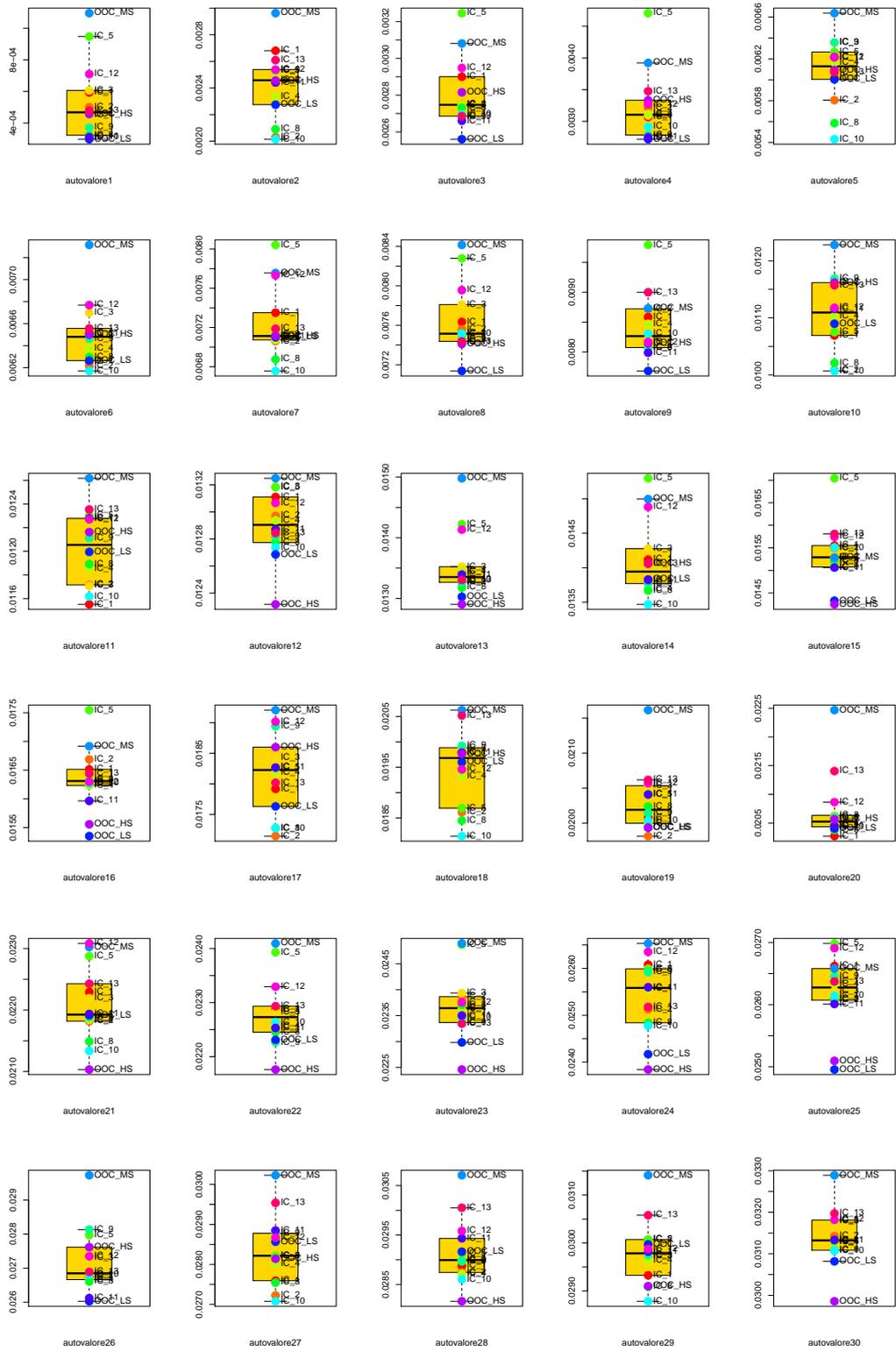


Fig.7.2 boxplot dei primi 30 autovalori per gli elementi del dataset

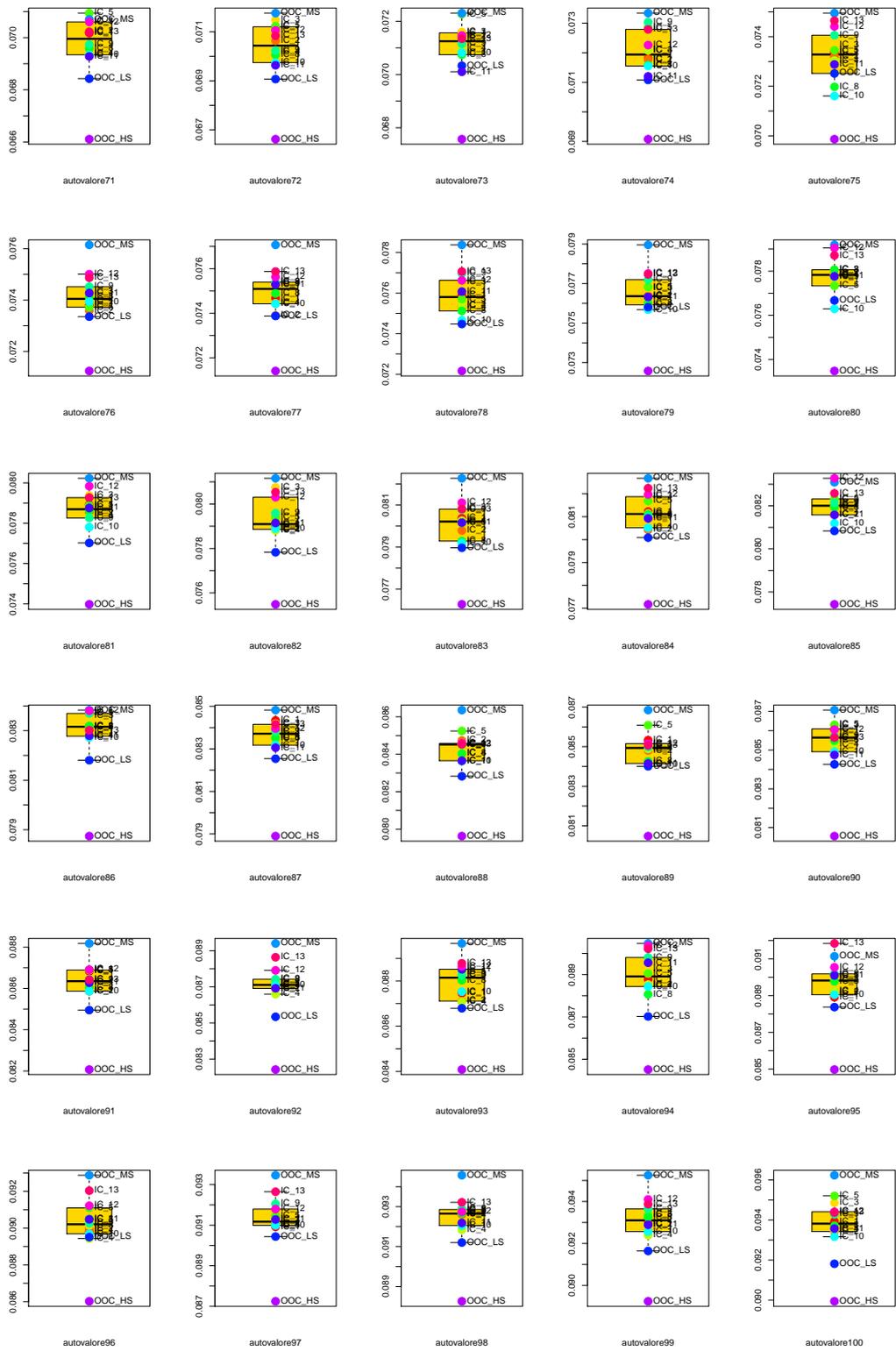


Fig.7.3 boxplot degli autovalori da 70 a 100

Capitolo 8

Conclusioni e sviluppi futuri

In questo lavoro sono stati presentati due metodi per l'analisi della variabilità in dataset di mesh, il primo è fondato sul concetto di distanza di Hausdorff e sulle tecniche dell'analisi di dati funzionali e composizionali, mentre il secondo ha per fulcro diversi risultati ottenuti nel corso degli anni nell'ambito della geometria differenziale e delle equazioni alle derivate parziali. Durante l'esposizione sono state avanzate numerose osservazioni riguardo i pregi, i difetti e le caratteristiche di entrambi i metodi, ma è opportuno ripetere brevemente ed arricchire tali osservazioni, per acquisire una piena consapevolezza dei limiti e delle potenzialità delle tecniche che sono state illustrate. Concentriamoci, per iniziare, sul calcolo dello spettro delle mesh.

- L'efficacia del metodo è, almeno teoricamente, del tutto indipendente dal particolare problema applicativo a cui si lavora: nell'ipotesi che nel mondo fisico isospettralità e isometria siano equivalenti, lo spettro di una varietà è una ricchissima fonte di informazioni.
- Il metodo si basa su una rappresentazione *intrinseca* dei dati, che prescinde dall'esistenza di un prototipo o da fasi di allineamento manuale dei dati, caratteristica che lo rende un buon candidato all'automatizzazione.
- Per ottenere risultati sensati, è necessario che le mesh rispettino certi standard di qualità e che il metodo di calcolo degli autovalori sia adatto, a livello di stabilità, accuratezza numerica e complessità computazionale, al particolare caso di applicazione. Nel nostro caso studio, si è sperimentato un esempio di questo fatto, in quanto due degli elementi in controllo sono stati eliminati dall'analisi spettrale proprio per problemi legati alla qualità delle mesh.
- Interpretare geometricamente le differenze tra successioni di autovalori appartenenti a superfici diverse sembra essere molto arduo, se non impossibile

Possiamo affermare dunque che l'applicazione del metodo spettrale beneficerà senz'altro, in futuro, del miglioramento dei metodi di calcolo degli autovalori, dell'aumento della potenza computazionale, e della qualità degli scanner 3D, elementi che potrebbero renderlo applicabile a mesh sempre più estese e complesse. Il problema dell'interpretabilità, d'altro canto, risulta complesso: in generale, sembra difficile immaginare di identificare in maniera precisa anomalie geometriche sulla base delle sole successioni spettrali.

Passando al primo metodo, le seguenti caratteristiche sono facilmente individuabili.

- Si tratta di un metodo generale e flessibile, applicabile ad una classe molto ampia di dataset di mesh.
- Come già osservato, il metodo fornisce risultati che hanno una pronta interpretazione fisica, un elemento di grande importanza nel Controllo Statistico in contesti produttivi.
- È necessaria la presenza di un prototipo, a cui i dati vanno allineati: è una fase che richiede l'intervento umano, elemento che può avere un discreto impatto sulle possibilità di automatizzazione del controllo.
- È necessario ridurre le densità oggetto di analisi ad un supporto unico, con la minima perdita di informazioni, e va dunque valutata, caso per caso, la ragionevolezza di questa semplificazione.

Si vede come, pertanto, il campo di applicabilità del metodo potrà essere esteso da progressi nelle tecniche di allineamento spaziale, che tendano verso la completa assenza di fasi manuali, e dallo sviluppo di teorie matematiche che introducano spazi di densità a cui sia permesso di annullarsi su un insieme di misura positiva, rendendo dunque superflua la fase di riduzione dei supporti.

Per concludere, la SFPCA basata sulle mappe di distanza trova il suo più grande punto di forza in un'interpretabilità immediata in termini di difettosità degli oggetti analizzati. Si immagini di condurre l'intera analisi svolta nel Capitolo 5 senza avere alcuna conoscenza preliminare delle caratteristiche del dataset: si sarebbe comunque in grado di risalire non solo all'identificazione degli elementi difettosi, ma anche alla tipologia dei difetti stessi (assenza, abbondanza di materiale, o entrambi). D'altro canto, l'analisi spettrale delle mesh gode di una grande generalizzabilità, in quanto basata su una rappresentazione che, se calcolata con la dovuta accuratezza, può essere considerata a tutti gli effetti una firma della forma di una superficie.

Bibliografia

- Aitchison, John (feb. 1982). “The Statistical Analysis of Compositional Data”. In: *Journal of the Royal Statistical Society. Series B (Methodological)* 44, pp. 139–177.
- (1986). *The Statistical Analysis of Compositional Data*. Chapman & Hall, Ltd. London. ISBN: 0-412-28060-4.
- Aspert, N, D Santa-Cruz e Touradj Ebrahimi (gen. 2002). “Mesh: Measuring errors between surfaces using the hausdorff distance”. In: *IEEE International Conference in Multimedia and Expo (ICME)*, pp. 705–708.
- Babu, Jogesh G., Angelo J. Canty e Yogendra P. Chaubey (lug. 2002). “Application of Bernstein Polynomials for smooth estimation of a distribution and density function”. In: *Journal of Statistical Planning and Inference* 105, pp. 377–392. DOI: 10.1016/S0378-3758(01)00265-8.
- Barcelo-Vidal, Carles, Josep A. Martín-Fernandez e Vera Pawlowsky-Glahn (2001). “Mathematical Foundations of Compositional Data Analysis”. In: Belkin, Mikhail e Partha Niyogi (gen. 2008). “Towards a theoretical foundation for Laplacian-based manifold methods”. In: *J. Comput. Syst. Sci.* 74, pp. 1289–1308.
- Besl, Paul e H.D. McKay (mar. 1992). “A method for registration of 3-D shapes. IEEE Trans Pattern Anal Mach Intell”. In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 14, pp. 239–256. DOI: 10.1109/34.121791.
- Boogaart, Karl Gerard van den, Juan Jose Egozcue e Vera Pawlowsky-Glahn (lug. 2010). “Bayes linear spaces”. In: *SORT* 34.
- Canzani, Yaiza (2013). *Notes for Analysis on manifolds via the Laplacian*. Springer International Publishing. URL: <http://www.math.harvard.edu/~canzani/docs/Laplacian.pdf>.
- Chen, Nan, Xuemin Zi e Changliang Zou (mag. 2015). “A Distribution-Free Multivariate Control Chart”. In: *Technometrics* 58, pp. 00–00. DOI: 10.1080/00401706.2015.1049750.
- Cignoni, Paolo, Claudio Rocchini e Roberto Scopigno (giu. 1998). “METRO: Measuring error on simplified surfaces”. In: *Computer Graphics Forum* 17, pp. 167–174. DOI: 10.1111/1467-8659.00236.
- Colosimo, Bianca e Massimo Pacella (ott. 2007). “On the use of principal component analysis to identify systematic patterns in roundness profiles”.

- In: *Quality and Reliability Engineering International* 23, pp. 707–725.
DOI: 10.1002/qre.878.
- Courant, R. e D. Hilbert (1963). *Methods of Mathematical Physics*. v. 1. Interscience Publishers.
- Eaton, Morris L. (lug. 1983). *Multivariate statistics : a vector space approach*. 1^a ed. John Wiley e Sons. ISBN: 9780471027768.
- Egozcue, Juan Jose (ott. 2009). “Reply to “On the Harker Variation Diagrams; ...” by J.A. Cortés”. In: *Mathematical Geosciences* 41, pp. 829–834. DOI: 10.1007/11004-009-9238-0.
- Egozcue, Juan Jose, J L. Díaz-Barrero e Vera Pawlowsky-Glahn (lug. 2006). “Hilbert Space of Probability Density Functions Based on Aitchison Geometry”. In: *Acta Mathematica Sinica, English Series* 22, pp. 1175–1182. DOI: 10.1007/s10114-005-0678-2.
- Filzmoser, Peter, Karel Hron e Matthias Templ (2018). *Applied Compositional Data Analysis*. Springer. ISBN: 978-3-319-96420-1.
- Godinho, L. e J. Natário (2014). *An Introduction to Riemannian Geometry: With Applications to Mechanics and Relativity*. Universitext. Springer International Publishing. ISBN: 9783319086668. URL: <https://books.google.it/books?id=oV4qBAAQBAJ>.
- Gordon, Carolyn Sue, D. L. Webb e Scott A. Wolpert (1992). “Isospectral plane domains and surfaces via Riemannian orbifolds”. In: *Inventiones mathematicae* 110, pp. 1–22.
- Grasso, Marco e Bianca Colosimo (giu. 2017). “Statistical Process Monitoring of Additive Manufacturing via In-Situ Sensing”. In:
- Guthe, Michael, Pavel Borodin e Reinhard Klein (gen. 2005). “Fast and Accurate Hausdorff Distance Calculation between Meshes.” In: vol. 13, pp. 41–48.
- Henrikson, Jeff (1999). “Completeness and Total Boundedness of the Hausdorff Metric”. In:
- Horváth, Lajos e Piotr Kokoszka (2012). *Inference for Functional Data with Applications*. Springer Series in Statistics. Springer New York. ISBN: 9781461436553. URL: https://books.google.it/books?id=0VezLB_ZpYC.
- Hron, K, Alessandra Menafoglio, Matthias Templ, Klára Hrušová e P Filzmoser (lug. 2014). “Simplicial principal component analysis for density functions in Bayes spaces”. In: *Computational Statistics & Data Analysis* 94. DOI: 10.1016/j.csda.2015.07.007.
- Huang, Xiaohu, Qiang Zhou, Li Zeng e Xiaodong Li (ott. 2015). “Monitoring Spatial Uniformity of Particle Distributions in Manufacturing Processes Using the K Function”. In: *IEEE Transactions on Automation Science and Engineering* 14, pp. 1–11. DOI: 10.1109/TASE.2015.2479088.
- Johnson, R.A. e D.W. Wichern (2007). *Applied Multivariate Statistical Analysis*. Applied Multivariate Statistical Analysis. Pearson Prentice

- Hall. ISBN: 9780131877153. URL: <https://books.google.it/books?id=gFWcQgAACAAJ>.
- Kac, Marc (apr. 1966). “Can One Hear the Shape of a Drum?” In: *The American Mathematical Monthly* 73, pp. 1–23.
- Lee, John M. (1997). *Riemannian Manifolds: An Introduction to Curvature*. Graduate Texts in Mathematics. Springer New York. ISBN: 9780387982717. URL: <https://books.google.it/books?id=ZRQgH7FQafgC>.
- Levy, Bruno (lug. 2006). “Laplace-Beltrami Eigenfunctions Towards an Algorithm That Understands Geometry”. In: vol. 2006, pp. 13–13. ISBN: 0-7695-2591-1. DOI: 10.1109/SMI.2006.21.
- Li, Xinge, Guoliang Xu e Yongjie Zhang (set. 2015). “Localized discrete Laplace-Beltrami operator over triangular mesh”. In: *Computer Aided Geometric Design* 39. DOI: 10.1016/j.cagd.2015.09.001.
- Machalová, J, Karel Hron e Gianna Monti (dic. 2015). “Preprocessing of centred logratio transformed density functions using smoothing splines”. In: *Journal of Applied Statistics* 43. DOI: 10.1080/02664763.2015.1103706.
- McKean, Henry e I M. Singer (gen. 1967). “Curvature and the Eigenvalues of the Laplacian”. In: *Journal of Differential Geometry - J DIFFERENTIAL GEOM* 1. DOI: 10.1007/978-3-319-22237-0_6.
- Menafoglio, Alessandra, Marco Grasso, Piercesare Secchi e Bianca Colosimo (feb. 2018). “Profile Monitoring of Probability Density Functions via Simplicial Functional PCA With Application to Image Data”. In: *Technometrics*. DOI: 10.1080/00401706.2018.1437473.
- Minakshisundaram, S. (1953). *Eigenfunctions on Riemannian Manifolds*.
- Oniga, Ersilia e Constantin Chirila (gen. 2013). “Hausdorff Distance for the Differences Calculation Between 3D Surfaces”. In: *Journal of Geodesy and Cadastre RevCAD* 15, pp. 193–202.
- Patané, Giuseppe (ott. 2014). “Laplacian spectral distances and kernels on 3D shapes”. In: *Pattern Recognition Letters* 47, 102–110. DOI: 10.1016/j.patrec.2014.04.003.
- (lug. 2017). “An introduction to Laplacian spectral distances and kernels: theory, computation, and applications”. In: pp. 1–54. DOI: 10.1145/3084873.3084919.
- Pawlowsky-Glahn, Vera, Juan Jose Egozcue e K Van den Boogaart (giu. 2014). “Bayes Hilbert Spaces”. In: *Australia and New Zealand Journal of Statistics* 56, 171–194. DOI: 10.1111/anzs.12074.
- Pearson, Karl (1897). “Mathematical contributions to the theory of evolution. On a form of spurious correlation which may arise when indices are used in the measurement of organs”. In: pp. 489–502. DOI: 10.1098/rsp1.1896.0076.
- Protter, M.H. (1987). “Can one hear the shape of a drum? Revisited”. In: *Society for Industrial and Applied Mathematics*.

- Ramsay, James e B. W. Silverman (2005). *Functional Data Analysis*. Springer. ISBN: 978-0387-40080-8.
- Reuter, Martin (2010). “Hierarchical Shape Segmentation and Registration via Topological Features of Laplace-Beltrami Eigenfunctions”. In: *International Journal of Computer Vision*. DOI: 10.1007/s11263-009-0278-1. URL: <http://dx.doi.org/10.1007/s11263-009-0278-1>.
- Reuter, Martin, Niklas Peinecke e Franz-Erich Wolter (apr. 2006). “Laplace-Beltrami spectra as ‘Shape-DNA’ of surfaces and solids”. In: *Computer-Aided Design* 38, pp. 342–366. DOI: 10.1016/j.cad.2005.10.011.
- Reuter, Martin, Marc Niethammer, Franz-Erich Wolter, Sylvain Bouix e Martha Shenton (ott. 2007). “Global Medical Shape Analysis Using the Volumetric Laplace Spectrum”. In: *Lecture notes in computer science* 2007, pp. 417–426. DOI: 10.1109/CW.2007.42.
- Rudin, Walter (1991). *Functional Analysis*. International series in pure and applied mathematics. McGraw-Hill. ISBN: 9780070542365. URL: https://books.google.it/books?id=Sh_vAAAAMAAJ.
- Salsa, S. (2010). *Equazioni a derivate parziali: Metodi, modelli e applicazioni*. UNITEXT. Springer Milan. ISBN: 9788847016460. URL: <https://books.google.it/books?id=cG69bf6e7BkC>.
- Verdière, Yves Colin de (1987). “Construction de laplaciens dont une partie finie du spectre est donnée”. In: *Annales Scientifiques de l'É.N.S.*
- Wetzler, Aaron, Yonathan Aflalo, Anastasia Dubrovina e Ron Kimmel (mag. 2013). “The Laplace-Beltrami Operator: A Ubiquitous Tool for Image and Shape Processing”. In: vol. 7883. DOI: 10.1007/978-3-642-38294-9_26.
- Yan, Hao, Kamran Paynabar e Jianjun Shi (ott. 2015). “Anomaly Detection in Images with Smooth Background Via Smooth-Sparse Decomposition”. In: *Technometrics* 59, pp. 00–00. DOI: 10.1080/00401706.2015.1102764.
- Zhao, Xueqi e Enrique Del Castillo (giu. 2019). “An Intrinsic Geometrical Approach for Statistical Process Control of Surface and Manifold Data”. In: *ESAMLAB, Penn State*.